UNIVERSITY OF ALBERTA

TRANSCRIPTIONAL MAPPING IN THE PROXIMAL REGION OF HUMAN CHROMOSOME 22

BY

MOHAMMAD ALI RIAZI ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

in

Molecular Biology and Genetics

Department of Biological Sciences

Edmonton, Alberta
Spring, 1998

National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-29098-0

Canada

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled "Transcriptional mapping in the proximal region of human chromosome 22" submitted by Mohammad Ali Riazi in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Molecular Biology and Genetics.

Heather M McDermid
Dr. H.E. McDermid (supervisor)

Dr. D.B. Pilgrim

Roseline Godbout
Dr. R. Godbout

Dr. P. Wong

P.J. McAlpine
Dr. P.J. McAlpine

DATE: Jan 19, 1998

This thesis is dedicated to the memory of my father

Reza Riazi

Abstract

Human chromosome 22 cytogenetic band 22q11 is predicted to be a gene rich region as well as being the location of a number of rearrangements leading to syndromes such as cat eye syndrome (CES). CES is a developmental disorder characterized by ocular coloboma, imperforate anus, preauricular skin tags and pits, heart defects, kidney, mild mental retardation, and dysmorphic features such as hypertelorism and downslanting palpebral fissures. The syndrome is usually associated with the presence of a bisatellited supernumerary chromosome derived from an inverted duplication of 22pter-22q11.2. This results in four copies of the region being present in the patients. The minimal duplicated critical region is defined as proximal to locus D22S57, covering approximately 2 Mb.

The purpose of this study was to identify genes within the critical region of this syndrome using a variety of approaches, including "exon trapping" and genomic sequence analysis. This study has resulted in the identification of many putative genes. Some of these putative genes have been characterized and could have some significance in the etiology of CES. The gene "IDGFL" could function as a growth factor based on the similarity to a novel insect growth factor. A set of three possible alternatively spliced transcripts of the gene "SAHL1" were also characterized. SAHL1-1 is expressed in the fetal and adult kidney and liver, and therefore could play a part in some features of CES. This study also added to the growing number of unprocessed pseudogenes mapping to the proximal region of chromosome 22. The gene "KCNMBL", identified in this study, which seems to function as a regulatory subunit of Ca^{+2}-activated potassium channels, maps to 3q26.3-q27 (the region duplicated in the dup 3(q) syndrome) and has a related sequence in the critical region of CES, which is probably a truncated pseudogene. Further characterization of the putative genes identified within the critical region of CES will help in the elucidation of the molecular defects leading to the features of CES.

Acknowledgments

I would like to greatly appreciate Heather McDermid for all her help, patience, and encouragement throughout the course of this project. She was always willing to listen to my hypotheses no matter how unreal they sounded. I also thank my supervisory committee, Drs. D.B. Pilgrim and R. Godbout for their invaluable suggestions. Thanks to our US collaborators, Bev and Marcia for giving me a share of the CATCH22 project. Dr. Shimizu's lab (Japanese collaborators), specially Mino, was a great help and kindly provided me a partial genomic sequence of the region as well as the sequence of some of the cDNAs. Dr. D. Rancourt (University of Calgary), Judy (Animal Services, Dept. of Bio. Sci.), and Violet (Dept. of Psychology) are also appreciated for helping me to get started with the mouse genetics and molecular biology techniques. I am grateful to all the people in McDermid's lab for being perfect coworkers and getting along. Talking with them was always a source of pleasure.

Last but not the least, I would like to express my sincere thanks to my wife Azam for being patient, and supporting me for all these years, and for still believing in me as a good husband, and finally my little angels Mahdis and Mahta for still loving a busy dad.

Table of Contents

Chapter I: Introduction **Page**

Human chromosome 22 1
Anomalies associated with the cytogenetic band, 22q11 1
Cat eye syndrome 3
Syndromes showing overlapping phenotypes with CES 9
DiGeorge/Velocardiofacial syndrome ("CATCH22") 13
Aneuploidy: Shapiro and Epstein controversy 15
Defining the minimal critical region of cat eye syndrome 18
Human genome: gene frequency, distribution, and organization 22
22q11: gene-rich or gene-poor 24
Identification of genes by positional cloning 24
Genetic mapping, physical mapping, and cloning of chromosome 22 25
Transcript mapping 27
Approaches for identification of genes 28
  A. CpG island identification and cloning 28
  B. Exon trapping 29
  C. cDNA selection 30
  D. "UniGene" mapping 30
  E. Computer analysis of DNA sequences 31
  F. Comparative mapping 32
Future objectives 35

Chapter II: Materials and Methods

Tissue culture 36
Genomic DNA extraction 36
Plasmid and cosmid DNA preparation 37
Preparation of bacterial artificial chromosome (BAC) DNA 37
Preparation of yeast artificial chromosome (YAC) DNA and PFGE of YACs 38
Southern blotting and hybridization of DNA probes 38
Human and mouse cDNA and genomic library screening 39
  Plating the cDNA libraries 39

Plaque and colony lifting and hybridization 39
Screening a mouse BAC genomic library 41
Preparation of total and poly A^+ RNA 41
Northern blotting 42
Polymerase chain reaction 42
Reverse transcription-PCR (RT-PCR) 42
Rapid amplification of 3' end of cDNA (3' RACE) 45
Exon trapping 45
1. Subcloning into pSPL3 or pSPL3B 45
2. Transformation, RNA extraction, and reverse transcription 48
3. Polymerase chain reaction to amplify the putative exons 48
4. Confirming the origin of the exons 49
Sequencing of the exons and cDNA 49
Database searching 49
Mouse whole mount *in situ* hybridization 50
1. Collection and storage of embryos 50
2. Preparation of digoxigenin-labeled RNA probes 51
3. Incorporation efficiency 51
5. Hybridization of RNA probes to embryos 52
5. Removing nonspecific hybridization 52
6. Antibody detection and visualization of the signals 52
7. Preparation of acetone powder and subtraction of antibody 53
Fluorescence *in situ* hybridization 53
1. Preparation of metaphase cells and slides 53
2. Preparation of biotin-labeled probes 55
3. Hybridization, washing, and visualization of the signal 55

Chapter III: Results

Part A: Cloning within the CESCR and isolation of exons in the region 56
1. A YAC contig within the CESCR 56
2. Isolation and analysis of putative exons 56
Part B: Characterization of putative genes in the CESCR 64
1. Cloning and characterization of pseudogene KCNMBLP, in the vicinity of D22S795, and gene KCNMBL on chromosome 3 64
Exon isolation and 3' end cloning 64

Isolation of a cDNA by computer sequence search 64
Southern analysis of KCNMBL 67
Sequence comparison of KCNMBL with the sequence of the putative exons on b7G2 in the CESCR 67
Study of the expression of KCNMBL 74
2. Partial cloning and characterization of a gene, SAHL1, adjacent to D22S9 ...74
Northern hybridization of SAHL1-1 to human and mouse RNA 79
Identification of two cDNAs from a Caco-2 cDNA library 79
Analysis of the cosmid c1A3 sequence reveals partial genomic structure of SAHL1-1, 2, and 3 79
Mouse embryonic expression of SAHL1 84
Isolation of a mouse BAC for SAHL1 to search for a region of the mouse genome homologous to the CESCR 84
3. Cloning and characterization of a gene, IDGFL in the vicinity of D22S43 87
Identification of IDGFL cDNA and sequence analysis 87
Northern analysis of IDGFL 87
Attempts to isolate a mouse homologue for IDGFL 96
4. Characterization of e7 and e8 96
Isolation of the exons and cDNAs 96
Isolation of the mouse genomic and cDNA clones for e7 97
Search for candidate genes in the region 97
1. CES-TK 97
2. PAX6, pnut, VEGF 102
Part C: Molecular localization of translocations in proximal 22q11 105
Translocation mapping of TOH (GM01700) 105
Translocation mapping in GM07511A 112

Chapter IV: Discussion

Part A: Cloning within the CESCR 114
A contig of the CESCR composed of YACs, BACs, and PACs 114
Part B: Identification of genes in the CESCR 115
Amplification of putative gene segments 116
Identification of a novel gene, KCNMBL, on chromosome 3 and a related sequence from the CESCR 117

KCNMBL, a putative candidate for some features of dup (3q) syndrome ...119
Pericentric region of chromosome 22 is rich in unprocessed pseudogenes ...120
IDGFL, a putative candidate for some features of CES 123
A putative gene, SAHL1 125
Comparative mapping of the CESCR 126
Developmental defects in CES, a molecular etiology 127
Future research 128
Identification of genes from the CESCR 128
Identification of the region in mouse with conserved synteny to the CESCR ...129
Further characterization of IDGFL, SAHL1, and some candidate genes 129
Toward a mouse model for CES 130
Part C: Is there a translocation/rearrangement hot spot on 22q11? 130
A translocation t(21;22)(p12;q11.2) disrupts *CLTCL* gene in the CATCH22 region in TOH 131
IGL locus involved in translocation t(14;22)(q32;q11.2) in GM07511A 133

References 134

Appendix 152

List of Tables

Number	Title	Page
Table 1-1	Common defects in CES	5
Table 1-2	The syndromes with partial phenotypic similarity to CES	11
Table 2-1	Genomic and cDNA libraries used in this study	40
Table 2-2	List of the primers used in this study	44
Table 2-3	Translocation cell lines used in FISH analysis	54
Table 3-1	The putative exons amplified from the CESCR using exon trapping	63
Table 3-2	Translocation breakpoint mapping at chromosome 22q11	106

List of Figures

Number	Title	Page
Figure 1-1	An ideogram of chromosome 22	2
Figure 1-2	Models for formation of the CEC	8
Figure 1-3	A transcript map of the CATCH22 critical region	17
Figure 1-4	Cat eye syndrome critical region (CESCR)	21
Figure 1-5	A human/mouse comparative map of 22q11	34
Figure 2-1	Schematic representation of the exon trapping procedure	47
Figure 3-1	Sizing of YACs by PFGE	58
Figure 3-2	A YAC contig map of CESCR	59
Figure 3-3	A diagram of the clones used in exon trapping and the gene segments identified	62
Figure 3-4	KCNMBL and KCNMBLP genomic structure and sequence	66
Figure 3-5	The comparison of the deduced amino acid sequence of KCNMBL with CO6 and hslo-beta proteins	69
Figure 3-6	Southern analysis of KCNMBL	71
Figure 3-7	Sequence comparison between KCNMBL on chromosome 3 and KCNMBLP on chromosome 22	73
Figure 3-8	Expression of KCNMBL	76
Figure 3-9	Southern and Zoo blot analysis of e22	78
Figure 3-10	Expression of SAHL1 in human and mouse	81
Figure 3-11	Alternative splicing and partial genomic structure of SAHL1	83
Figure 3-12	Whole mount *in situ* hybridization of SAHL1	86
Figure 3-13	Southern hybridization of e5	89
Figure 3-14	Sequence of the IDGFL cDNA	91
Figure 3-15	Protein sequence comparison of IDGFL with IDGF and AGSA proteins	93
Figure 3-16	Expression analysis of IDGFL and the putative mouse homologue, mC1	95
Figure 3-17	The sequence of cDNA c7/8	99
Figure 3-18	Fish analysis of m7B	101
Figure 3-19	Human and mouse Northern analysis of CES-TK and mtk	104
Figure 3-20	Cosmid contig adjacent to D22S75 and FISH analysis of	

these cosmids in a patient (TOH) carrying a t(21;22)(p12;q11.2) ..109

Figure 3-21 FISH analysis of the cosmids in the D22S75 region hybridized to metaphase spreads from TOH 111

Figure 3-22 A partial map of the IGL locus on chromosome 22 113

Abbreviations

AWS: Alagille-Watson syndrome
BAC: bacterial artificial chromosome
BD: branchiootorenal dysplasia
BGS: Baller-Gerold syndrome
bp: base pair
KCNMBL: Ca^{+2}-activated potassium channel beta subunit like
KCNMBLP: Ca^{+2}-activated potassium channel beta subunit like pseudogene
CATCH22: cardiac defect, abnormal facies, thymic hypoplasia, cleft palate, and hypocalcemia on chromosome 22
CEC: cat eye chromosome
CEN: centromere
CES: cat eye syndrome
CESCR: cat eye syndrome critical region
CHARGE: CHARGE association
DGS: DiGeorge syndrome
FISH: Fluorescence in situ hybridization
FTAPVR: familial total anomalous pulmonary venous return
FTOF: familial tetralogy of Fallot
IDGFL: insect-derived growth factor like
kb: kilobase pairs
MAA: microphtalmia or anophthalmos with associated anomalies
Mb: megabase pairs
MKS: Meckel syndrome
ONCR: optic nerve coloboma with renal disease
Opitz: Opitz GBBB
ORF: open reading frame
PAC: P1 artificial chromosome
PBS: prune belly syndrome
PCR: polymerase chain reaction
PFGE: pulsed field gel electrophoresis
PHS: Pallister-Hall syndrome
RACE: Rapid amplification of cDNA ends
RLMC: Rutledge-lethal multiple congenital anomaly

RT: room temperature
RT-PCR: reverse transcriptase PCR
SAHL1: SAH like 1
TBS: Townes-Brocks syndrome
TEL: telomere
UTR: untranslated region
VCFS: velocardiofacial syndrome
YAC: yeast artificial chromosome

Chapter I: Introduction

Human chromosome 22

Human chromosome 22 is an acrocentric chromosome with an estimated size of 53 Mb (Sanger Center World Wide Web page URL: www.sanger.ac.uk), which constitutes approximately 1.9% of the haploid autosomal genome (Morton, 1991). The p arm of this chromosome consists mostly of heterochromatin and ribosomal DNA as in the other acrocentrics (Schmickel et al., 1977). The longer q arm forms three major cytogenetic bands and ten subbands (figure 1-1) when Giemsa banded (G-banded), from the centromere, q11.1, to telomere, q13.3 (Harden and Klinger ISCN, 1985). The q arm is mostly euchromatin and has a high gene density (Saccone et al., 1996).

Due to its small size, and the localization of many syndrome and cancer-related rearrangements, chromosome 22 has been the focus of intensive mapping and sequencing efforts (Sanger Center WWW page). Currently the majority of the chromosome has been cloned into vectors suitable for sequencing, which is now underway (Kim et al., 1996). This information will be applied to the identification of many disease-associated genes on chromosome 22.

Anomalies associated with the cytogenetic band, 22q11

The cytogenetic band 22q11 is the site for a number of recurrent constitutional and tumor-associated rearrangements (Griffin et al., 1986). Two relatively frequent cancer-related translocations, associated with chronic myelogenous leukemia (CML) and Burkitt lymphoma, occur in 22q11, creating a chimeric gene (figure 1-1). In CML, a translocation t(9;22)(q34;q11) juxtaposes the c-*abl* protooncogene sequences on chromosome 9 with the 5' portion of the *bcr* gene on chromosome 22 (Groffen et al., 1994). In about 15% of Burkitt's lymphoma cases a translocation [t(8;22)(q24;q11)] causes the juxtaposition of the c-*myc* protooncogene and the immunoglobin λ light chain region on chromosome 22, causing deregulation of c-*myc* (Haluska et al., 1987).

In addition to cancer-related rearrangements, 22q11 is also the site of the most frequent non-Robertsonian congenital translocation t(11;22)(q23.3;q11.2). The translocation was described in more than 150 independent families prior to the 1980s (Fraccaro et al. 1980, Zackai and Emanuel, 1980). The translocation between the long arms of chromosome 11 and 22 (in the 22q11 cytogenetic band) is also found in association with peripheral neuroepitheliomas (NE) (Aurias et al., 1984; Whang-Peng et al., 1984). The

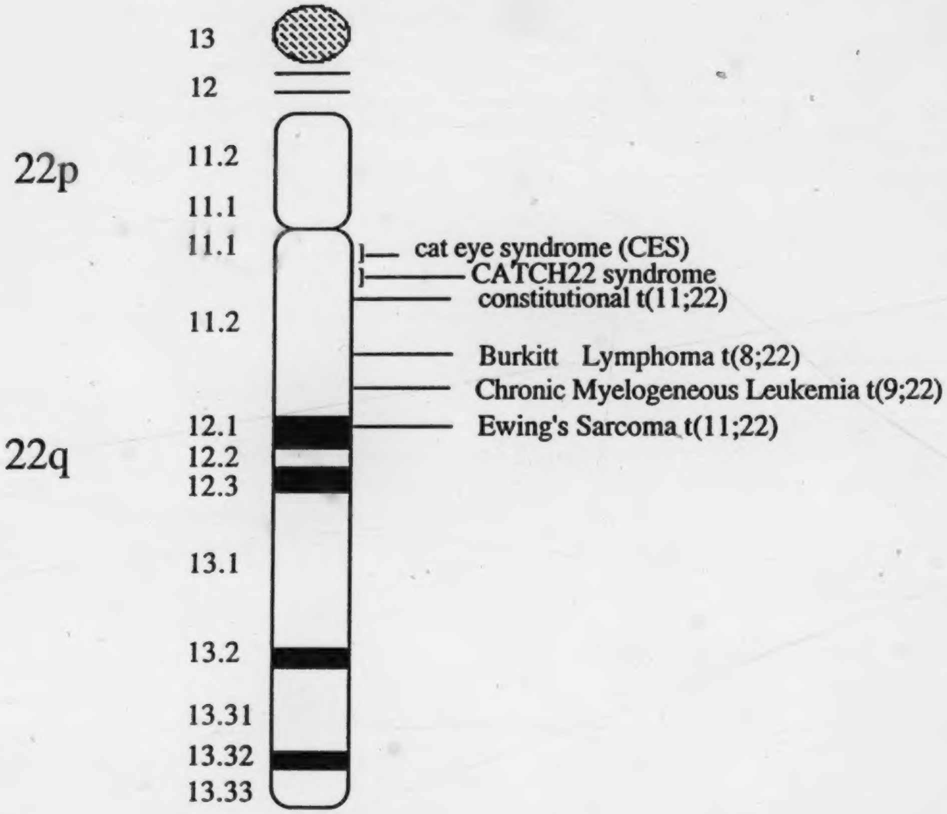

22p
22q
13
12
11.2
11.1
11.1
11.2
12.1
12.2
12.3
13.1
13.2
13.31
13.32
13.33
cat eye syndrome (CES)
CATCH22 syndrome
constitutional t(11;22)
Burkitt Lymphoma t(8;22)
Chronic Myelogeneous Leukemia t(9;22)
Ewing's Sarcoma t(11;22)

Figure 1-1: An ideogram of chromosome 22 (Harden and Klinger ISCN 1985). Cytogenetic band 22q11 (q11.1 + q11.2) is the site for many congenital and cancer-related rearrangements. Duplication of the most proximal region causes cat eye syndrome (CES). CATCH22 is caused by deletion of a region distal to CES.

translocation involved in Ewing sarcoma (ES) occurs in 22q12 (figure 1-1). Although both the constitutional and the tumor-associated t(11;22) originally seemed to localize to the same region by cytogenetic analysis, molecular analysis using DNA probes which map to proximal 22q and distal 11q showed that the translocation breakpoints were not in the same region on either chromosome (Budarf et al. 1989). Two other types of congenital rearrangements also occur on 22q11: CATCH22, and cat eye syndrome (CES) (figure 1-1). The former is usually caused by an interstitial deletion within 22q11, and the latter is the result of a duplication of the proximal region of 22q11. The chromosomally unbalanced offspring of t(11;22) carriers, who result from a 3:1 meiotic segregation, always carry a supernumerary der(22) t(11;22) and are therefore partially trisomic for chromosomes 11 and 22. The phenotype of these individuals overlaps with the phenotype of CES (Fraccaro et al., 1980).

The molecular basis for the frequency of chromosomal abnormalities on 22q11 is not understood; however it has been hypothesized to be the consequence of the presence of multiple recombinogenic low-copy repetitive sequences and gene families throughout the region (Lindsay et al., 1996; Halford et al, 1993). There have been reports of a number of chromosome specific repetitive sequences in 22q11, such as the KI-386 family (Carey et al., 1990), the gamma-glutamyl transferase (GGT) gene family (Bulle et al., 1987; Collins et al., 1997), and sequences related to the *bcr* gene (Heisterkamp and Groffen, 1988). There is evidence from other regions of the genome which supports this hypothesis. For example, the CRI-S232 and G1.3 repeat families are dispersed through Xq22.3 causing frequent deletions in STS deficiency (Yen et al., 1990), and Kallman syndrome (Schnur et al., 1990). The association of rearrangements and repetitive sequences has also been suggested for the proximal region of chromosome 15. The unequal sister chromatid exchange or looping out of DNA mediated by these repeats have been suggested to produce the deletions of 15q11.2. As well, the homologous pairing between these repeats in a U-type crossover (discussed below) is suggested to be the underlying cause of inv dup(15) (Donlon et al., 1986)

Cat eye syndrome

The association of coloboma (the lack of closure of the choroid fissure) and imperforate anus (the lack of rupture of anal membrane) was first noted over a century ago (Haab, 1878). These clinical manifestations, together with other features, are characteristics of cat eye syndrome, a term first used by Gerald et al., (1968). This is a rare developmental defect with an estimated incidence of one in approximately 50,000 to 150,000 (MIM

#115470) . The main abnormalities seen in these patients have been defined in table 1-1. These features are variable even within a single family, ranging from full CES phenotype to a near normal phenotype (Schinzel et al., 1981). The main features include: preauricular skin tags and pits; imperforate anus; unilateral or bilateral coloboma of the iris, choroid, and/or retina; congenital heart malformations, particularly total anomalous pulmonary venus return (TAPVR) and tetralogy of Fallot (TOF); renal malformations, particularly absence of one or both kidneys, and renal hypoplasia; and mild to moderate mental retardation (Schinzel et al., 1981).

Schachenmann et al. (1965) first noticed the presence of a supernumerary chromosome in the karyotypes of the patients with coloboma and imperforate anus. The "cat eye chromosome" or CEC (Buhler et al., 1972) is a dicentric bisatellited chromosome derived from an inverted duplication of the short arm and proximal long arm of chromosome 22 (inv dup 22pter-22q11.2) (Schinzel et al. 1981, McDermid et al., 1986). Although CECs are usually stable (Schinzel et al., 1981), there have been many cases of mosaicism where CECs are only found in a percentage of cells in the patients (Gerald et al., 1972; Hsu and Hirschhorn, 1977). Although the structure of CECs are usually stable, Urioste et al. (1994) reported a large variation in the size and shape as well as mosaicism of CECs in two sibs who inherited their CECs from their mother. This suggests that the CEC could undergo some structural changes during cell division. The occasional instability of the CECs is hypothesized to be the result of the dicentric nature of this chromosome. Normally, one of the centromeres is presumably inactivated, causing the stable transmission to both daughter cells during mitosis (Schreck et al., 1977). However, two functional centromeres in a dicentric chromosome such as CEC has been postulated to cause the loss or breakage and unequal exchanges in these chromosomes resulting in mosacisim in the number or the structure of CECs (Urioste et al., 1994). This is of clinical significance since a number of CES patients, who do not demonstrate any visible chromosome 22 abnormality, could have carried the CECs in their cells during embryonic development and then lost it in a number of tissues later.

The original cytogenetic studies of patients carrying a CEC suggested that this extra chromosome is derived from one of the acrocentric chromosomes because of the identification of satellites on both ends (Toomey et al., 1977). Due to its small size, chromosome banding studies could not clearly demonstrate the origin of the CEC and suggested chromosome 7 (Noel et al., 1973), 13 (Krmpotic et al., 1971), 14 (Pfeiffer et al., 1970), or 22 (Buhler et al., 1972; Schinzel et al., 1981) as its origin. Guanti et al. (1981) proposed that an interstitial deletion of chromosome 13 or a translocation between chromosome 13 and a G-group (21, 22) chromosome (probably chromosome 22) produces

Table 1-1: Common defects in CES. The definition of the phenotypes were obtained from Stedman's Medical Dictionary.

Phenotype	Definition
iris coloboma	congenital cleft of the iris, often associated with coloboma of the choroid
imperforate anus	absence of the normal opening of the anus
preauricular skin tags and pits	the outgrowth of epidermal and dermal fibrovascular tissue, and indentations anterior to the auricle of the ear
absent or hypoplastic kidney	absence or underdevelopment of kidney
total anomalous pulmonary venus return	failure of the connection of the pulmonary veins to the left atrium
tetralogy of Fallot	congenital heart disease consisting of high pulmonic stenosis, ventricular septal defect, dextroposition of aorta, and right ventricular hypertrophy

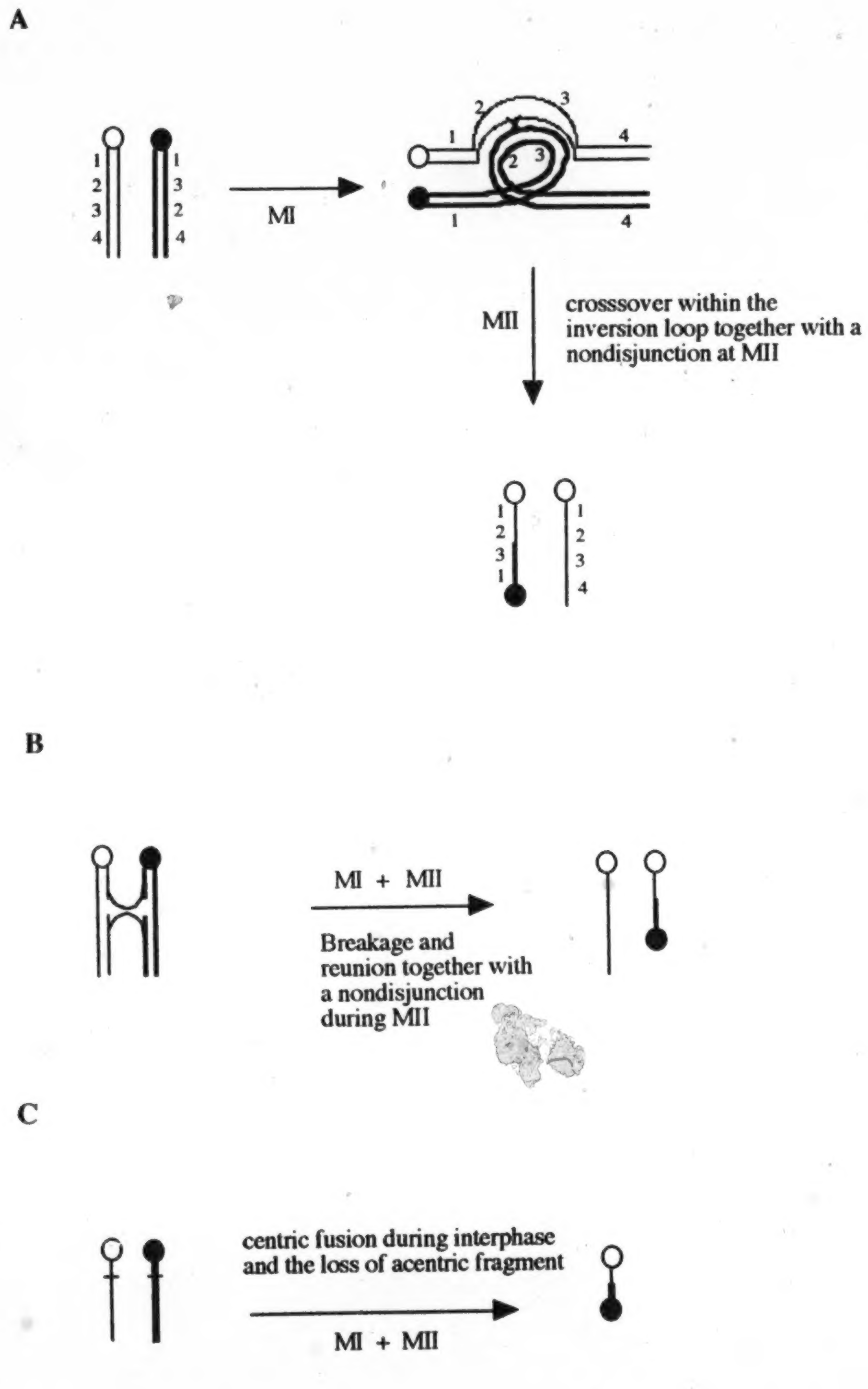
the CEC, based on the phenotypic similarities between trisomy 13 and cat eye syndrome patients. Finally, McDermid et al. (1986) confirmed the chromosome 22 origin of CEC by mapping a chromosome 22-derived single copy locus (D22S9) to the marker chromosome and showing that this locus was present in four copies in six cat eye syndrome patients carrying CECs.

Marker chromosomes, or abnormal chromosomes of undetermined origin, have also been observed in association with other abnormalities as well as in phenotypically normal individuals (Buckton 1985). Marker chromosomes derived from the proximal 15q (inv dup (15)) are suggested to account for about half of all the marker chromosomes with the rest being derived from a variety of other chromosomes (reviewed in Buckton 1985). The chromosome 15-derived markers demonstrate significant phenotypic heterogeneity. Although they have been found in phenotypically normal individuals, they have also been reported in association with mental retardation and Prader-Willi syndrome (Brondum-Nielson and Mikkelson 1995, Webb 1994).

The molecular mechanism which generates a chromosome such as the CEC or inv dup(15) is not understood but several hypotheses have been suggested. One hypothesis involves crossing over within an inversion loop present during meiosis in individuals heterozygous for a paracentric inversion followed by a nondisjunction at meiosis II. This would result in a gamete carrying a normal as well as a dicentric chromosome (Schreck et al. 1977). Since chromosome 22 is pale and small when G-banded, such an inversion would be difficult to detect. A second model to explain the derivation of the marker chromosomes proposes a U type exchange with breakage and abnormal reunion of the two non-sister chromatids during first meiotic division resulting in a dicentric and an acentric fragment. The acentric fragment is later lost while the dicentric one is retained in the gamete (Schreck et al. 1977, Van Dyke et al. 1977). Wisniewski et al. (1979) favored the latter model based mainly on their observation of the polymorphic satellites at the two ends of the inv dup (15) which they studied. This suggested that the bisatellited marker chromosomes were derived from a U type exchange between the two homologues. A third hypothesis proposes a centric fusion between the two homologues in one parent and a nondisjunction of the two normal chromosomes in the other parent to produce a proband with two normal chromosome and a supernumerary bisatellited chromosome (reviewed in Van Dyke et al., 1977, figure 1-2).

The majority of inv dup(15) chromosomes studied to date show a maternal origin (Crolla et al., 1994). Similarly Magenis et al. (1988) determined the maternal origin of the CEC in two patients by comparing the CEC Q-banded p-arm satellites with the chromosome 22 satellites in the parents. Using restriction fragment length polymorphism

Figure 1-2: Models for the formation of the CEC. A: crossing over in the inversion loop of a carrier for a pericentric inversion followed by a nondisjunction during meiosis II (MII) produces gametes with a normal and a dicentric chromosome. B: The breakage and reunion of the nonsister chromatids during meiosis I together with a nondisjunction during MII produces a gamete carrying a dicentric chromosme. C: a centric fusion results in a gamete contaianing a dicentric chromosome. Model C can explain the production of CES patients only if a nondisjunction occurs in the other parent, resulting in two normal chromosome 22 and a dicentric chromosome.


A
MI
MII
crosssover within the inversion loop together with a nondisjunction at MII
B
MI + MII
Breakage and reunion together with a nondisjunction during MII
C
centric fusion during interphase and the loss of acentric fragment
MI + MII

(RFLP) analysis of polymorphic markers, Mears (PhD thesis, 1995), described two individuals with cat eye syndrome who inherited their CEC from their mother. The possibility of transmission of the CEC from the father has also been documented (Noel et al., 1976). Mears et al. (1995) reported a family in which a ring chromosome derived from the proximal region of chromosome 22 (r(22)) was transmitted paternally over two generations causing all the features of CES in the proband.

The maternal as well as paternal transmission of CEC or r(22) in CES patients suggests the absence of imprinted genes, causing an abnormal phenotype, in this region. In addition, the apparent normal phenotype of an individual carrying a balanced Robertsonian t(22q;22q) inherited from his mother suggests the lack of maternally imprinted genes with major effect on chromosome 22 (Schinzel et al., 1994).

Syndromes showing overlapping phenotypes with CES

A number of patients who are diagnosed with CES do not carry a CEC or other visible chromosome 22 abnormality. Although the phenotype in these patients could be due to a small rearrangement which is cytogenetically undetectable, the abnormal features may have a different etiology. Individual and combinations of features of cat eye syndrome are also seen in association with other human syndromes. For example, imperforate anus is a feature of Townes-Brock syndrome (Townes and Brocks, 1972), Opitz GBBB, as well as some of the other syndromes (MIM database, table 1-2). Townes-Brock syndrome (TBS) in particular shows considerable overlap with CES. TBS is a developmental defect with abnormalities such as imperforate anus, anomalies of the hands and feet, auricular pits, fistulas, and tags, hypoplastic kidneys, and congenital heart defects (Serville et al., 1993). However, Friedman (1987) described a father and daughter with TBS, having a pericentric inversion of proximal region of chromosome 16. Serville (1993) suggested the gene for TBS might be in 16q12.1 after describing a TBS patient with a balanced translocation between chromosomes 5 and 16q12.1.

The association of coloboma of the optic nerve and renal hypoplasia have been recently defined as a separate syndrome, renal-coloboma syndrome (Sanyanusin et al. 1995). Both renal defects and optic nerve coloboma has been observed in the cat eye syndrome patients (MIM, # 115470). The renal-coloboma syndrome is now known to be caused by mutations in the *PAX2* gene. Sanyansuin et al (1995) predicted that patients having eye and renal defects could carry mutations in the *PAX2* gene on the basis of the expression pattern of this gene. *Pax2* in mice is expressed in the optic and otic vesicles, the mesonephros (which later gives rise to male and female genital tracts), kidney, spinal cord,

Table 1-2: The syndromes with partial phenotypic similarity to CES. This table demonstrates some of the syndromes listed in the OMIM database with at least two features of CES. The most common abnormalities of CES includes: IA (imperforate anus), CO (coloboma), PTP (preauricular tags and pits), CHD (congenital heart disease, TAPVR, TOF), RA (renal anomaly). Followings are the name of these syndromes and their map locations if known: **AWS**: Alagille-Watson syndrome; 20p11.2 (MIM #118450), **BD**: branchiootorenal dysplasia; 8q13.3 (MIM #113650), **BGS**: Baller-Gerold syndrome (MIM #218600), **CES**: cat eye syndrome; 22q11.2 (MIM #115470), **CHARGE**: CHARGE association (MIM #214800), **DGS**: DiGeorge syndrome; 22q11.2 (MIM #188400), **DS**: Down syndrome; 21q22.3 (MIM #190685), **FTAPVR**: familial total anomalous pulmonary venous return; 4p13-q12 (MIM #106700), **FTOF**: familial tetralogy of Fallot (MIM #187500), **JS**: Joubert syndrome (MIM #243910), **MAA**: microphthalmia or anophthalmos, with associated anomalies (MIM #309800), **MKS**: Meckel syndrome; 17q21-q24 (MIM #249000), **ONCR**: optic nerve coloboma with renal disease; 10q24.3-q25-1 (MIM #120330), **Opitz**: Opitz GBBB; 22q11.2 (MIM #145410), **PBS**: prune belly syndrome (MIM #100100), **PKHD1**: polycyctic kidney and hepatic disease; 6p21.1-p12 (MIM #263200), **PHS**: Pallister-Hall syndrome; 7p13 (MIM #146510), **RLMC**: Rutledge lethal multiple congenital anomaly syndrome (MIM #268670), **TBS**: Townes-Brocks syndrome; 16p12 (MIM #107480), **UMS**: ulnar-mammary syndrome; 12q23-q24.1 (MIM #181450), **VATER**: Vacteral association (MIM #192350), **ZNS**: Zunich neuroectodermal syndrome (MIM #280000). Candidate genes have only been identified for ONCR and BD. Mutations in *PAX2* on 10q24.3-q25.1 and EYA1 on 8q13.3 cause the defects in ONCR and BD respectively. *EYA1* is the human homolog of the "eyes absent" gene in *Drosophila*. (MIM #601653).

	IA	CO	PTP	CHD	RA
IA		CES, OPITZ, MAA	CES, TBS	CES, TBS, OPITZ, DS, VATER, AWS, UMS,	CES, TBS, OPITZ, PHS, PBS, VATER, AWS, BGS
CO			CES	CES, OPITZ, CHARGE, ZNS	CES, OPITZ, ONCR, MKS, JS, CHARGE
PTP				CES, TBS, FTOF	CES, TBS, BD
CHD				FTAPVR	CES, TBS, OPITZ, PKHD1, DGS, RLMC, AWS
RA					

midbrain, and hindbrain (Dressler et al., 1990). Krd mutant mice lacking a portion of chromosome 19 including the *Pax2* gene have reduced thickness of the renal cortex, a reduced number of glomeruli at birth, and reduced amplitudes on electroretinogram (Keller et al., 1994). In addition, the constitutive expression of *Pax2* in transgenic mice results in polycystic kidney abnormalities during gestation (Dressler et al., 1993). The association of *PAX2* and the renal-coloboma syndrome was demonstrated after a frameshift mutation producing a presumably truncated *PAX2* protein was discovered in a family with optic nerve coloboma and renal hypoplasia (Sanyanusin, 1995). In another study, Schimmenti et al. (1997) screened patients with variable disease associations of eye, renal, CNS, and hearing abnormalities for *PAX2* mutations. The eye abnormalities included optic nerve hypoplasia or aplasia, and bilateral retinal or optic nerve coloboma. They discovered mutations in exon 2 of the *PAX2* gene in 10% of the patients whom they studied. Interestingly one of the patient with mutated *PAX2* was originally diagnosed as having CHARGE association.

CHARGE association is probably a multifactorial disorder with unknown etiology and showing coloboma of the eye, heart anomaly, choanal atresia, mental retardation, genital anomalies, and ear abnormalities (MIM # 214800). Emanuel et al. (1992), reported a 22q11 deletion in 1/18 of the CHARGE patients they studied. CHARGE anomaly has been also seen in association with duplication of chromosome 14 (q22-q24.3), as well as being X-linked (North et al., 1995; Abruzzo and Erickson, 1989).

Total anomalous pulmonary venous return (TAPVR), the most common heart defect in CES, has been reported as both an isolated cardiac abnormality and a familial defect. TAPVR is a congenital heart defect in which the pulmonary veins fail to connect to the left atrium during cardiac development and instead they connect to the right atrium or one of its venous tributaries. A familial example was reported by Blyel et al. (1994), in a large Utah-Idaho kindred in which 14 individuals related over eight generations had this defect. Linkage analysis using polymorphic markers suggested localization of the gene(s) to approximately 30 cM at 4p13-q12. Several chromosome 22q11 markers were used in this study and showed no sign of linkage to this disease (Blyel et al., 1995). A number of genes have been identified in the 30 cM interval on 4p13-q12. Among these genes, a tyrosine kinase receptor gene (KDR, Terman et al. 1991) which carries a domain known as kinase insert domain could be a likely candidate for the gene involved in familial TAPVR. KDR and its mouse homologue, fetal liver kinase (FLK1, Mathews et al., 1991) bind vascular endothelial growth factor (VEGF) with high affinity *in vitro*, and thus are involved in a signaling pathway important in the development of vasculature (Terman et al., 1991).

Table 1-2 summarizes a number of the known syndromes which share some of the features seen in CES. These syndromes affect the same organs and produce a similar phenotype to CES, therefore the genes involved in these syndromes may affect the same molecular pathways leading to these defects. Consequently, the identification of genes associated with these syndromes might help in the elucidation of the molecular etiology of CES.

DiGeorge/Velocardiofacial syndrome ("CATCH22")

The original cytogenetic studies mapped the CES critical region to the same region as a deletion syndrome called DiGeorge syndrome (DGS). Consequently, it was hypothesized that both CES and DGS shared the same critical region, such that CES was caused by duplication of the region deleted in DGS (Sharkey et al., 1992). However, using dosage analysis of the probes clearly mapping to either CES or DGS critical regions, Mears et al. (1994) showed that the critical region for these two syndromes were different with CES mapping proximal to DGS (figure 1-1).

DGS is a developmental field defect of the third and fourth pharyngeal pouches (Lammer and Opitz, 1986) and maps to 22q11.2, distal to the critical region for cat eye syndrome (Mears et al. 1994). DGS is characterized by the absence or hypoplasia of the thymus and parathyroid glands, cardiac malformations, and dysmorphic facial features in some patients (DiGeorge, 1965; Conley et al., 1979). The cardiac anomalies together with occasional hypocalcemia, hypoplastic or absent lymphoid tissue, and T-cell deficiency show overlap with features of another syndrome called velocardiofacial syndrome (VCFS), suggesting the same or overlapping genetic etiology in these two disorders (Stevens et al. 1990). Velocardiofacial syndrome (VCFS) is an autosomal dominant disorder characterized by cleft palate, characteristic facial features, cardiovascular anomalies, and learning difficulties (Strong, 1968; Kinouchi et al., 1976; Shprintzen et al., 1978).

Cytogenetic studies have revealed that 15-20% of the patients with DGS have either an unbalanced translocation leading to monosomy 22pter-q11 (de la Chapelle et al., 1981; Kelly et al., 1982; Greenberg et al., 1988) or a visible interstitial deletion, del(22)(q11.21q11.23) (Greenberg et al., 1988; Mascarello et al., 1989; Driscoll et al., 1992a). RFLP, dosage and in situ hybridization analysis have demonstrated that in the majority of cases of DGS where no visible cytogenetic deletion was observed, a submicroscopic deletion within 22q11 could be detected (Driscoll et al., 1992a; Scambler et al., 1991). Similar deletions of 22q11.2 were found in as many as 90% of VCFS patients (Driscoll et al., 1992b; Carey et al., 1992).

Hemizygosity of this region has also been demonstrated in patients with conotruncal anomaly face (CTAF) syndrome, Kousseff syndrome, Cayler syndrome, Opitz GBBB syndrome, and absent-pulmonary-valve syndrome (reviewed in Dallapiccola et al. 1996). The overlapping features of these syndromes has led to the suggestion that these syndromes represent an abnormal clinical spectrum resulting from the deletion of the same region on chromosome 22. Consequently the acronym CATCH22 (cardiac defects, abnormal facial features, thymic hypoplsia, cleft palate, and hypocalcemia on chromosome 22) has been proposed to encompass the various clinical manifestations in these syndromes (Wilson et al., 1993). CATCH22 is known to be a major cause of developmental heart defects, second only to Down's syndrome (Burn and Goodship, 1996). The minimal critical region for CATCH22 is approximately 250 kb which is confined on the centromeric side by the common proximal deletion endpoint (proximal to D22S75) and distally by the breakpoint of an unbalanced t(15;22) in a patient with mild features of CATCH22 (Jaquez et al., 1997).

Although CATCH22 is usually associated with monosomy of the 22q11.2 region, a single case associated with a balanced translocation has been reported. The patient (ADU) has incomplete DGS, having telecanthus, microretrognathia, severe aortic coarcation with hypoplastic left aortic arch, decreased E rosettes and mild neonatal hypocalcemia. The patient carries a balanced translocation, t(2;22)(q14;q11.1) (Augusseau et al., 1986). The same translocation is also present in the mother of the proband, VDU, causing a milder phenotype which most closely resembles mild VCFS (Demczuk et al., 1995). The translocation breakpoint in this family has been localized and mapped to the CATCH22 critical region (Budarf et al., 1995a; Demczuk et al., 1995; Wadey et al., 1995). The association of balanced translocations with human syndromes and the subsequent cloning of the breakpoint has helped substantially in the discovery of the genes underlying different human syndromes (Mannens et al., 1996; Ross et al., 1997). The cloning of the ADU breakpoint has identified two potential open reading frame spanning the translocation breakpoint: DGCR3, potentially encoding a 260 amino acid protein with weak similarity to a region of the androgen receptor, and DGCR4 which is a putative gene with a 1.6 kb transcript and no similarity to previously identified genes (Budarf et al., 1995a). Sutherland et al. (1996) isolated several cDNAs using two cosmids and a fosmid clone encompassing the ADU breakpoint. The gene represented by these cDNAs, DGCR5 is also disrupted by ADU translocation breakpoint. Despite being transcribed and alternatively spliced, DGCR5 does not have any apparent protein product (Sutherland et al. 1996). It has been suggested that the haploinsufficiency of DGCR3, DGCR4, and DGCR5 might be responsible for the major features of CATCH22. However, attempts to findmutations in these genes in the

CATCH22 patients with no detectable deletion of 22q11 have so far failed (Dallapiccola et al., 1996). In addition a patient with very mild features of DGS, patient G, has been described who has a 22q11 deletion not extending into the ADU breakpoint region (Levy et al., 1995). This patient does not have a full CATCH22 phenotype, and therefore, his deletion may not encompass all genes involved in CATCH22 (if this syndrome is caused by more than one gene). Alternatively, the ADU breakpoint could disrupt the normal regulation of genes proximal or distal to the translocation via position effect (Budarf et al., 1995a).

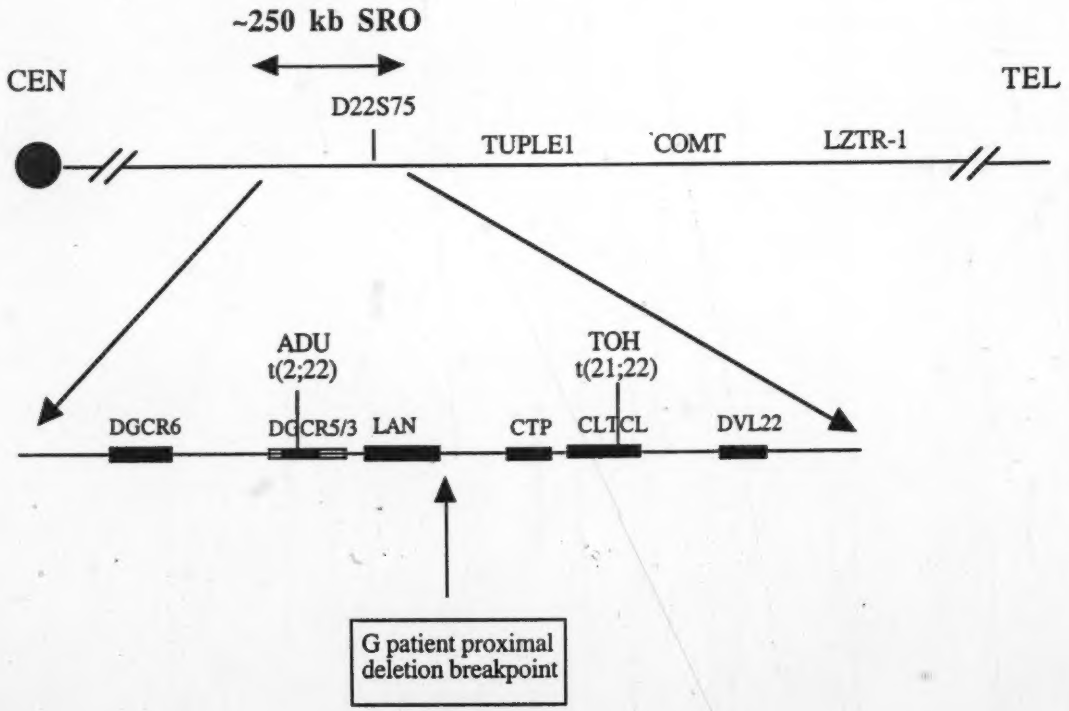
In addition to the genes DGCR3, DGCR4, and DGCR5, a number of other genes have also been localized to the critical region (figure 1-3). DGCR2/IDD/LAN which encodes a transmembrane protein, maps approximately 10 kb telomeric to the ADU breakpoint (Budarf et al. 1995a, Demczuk et al. 1995, Wadey et al. 1995). A transcriptional regulator protein called DGCR1/HIRA/TUPLE also localizes to about 180 kb telomeric from the ADU breakpoint (Halford et al. 1993, Lamour et al. 1995). Other genes mapping to the critical region are clathrin heavy chain-like gene (*CLTCL*, Long et al., 1996), mitochondrial citrate-transporter protein (*CTP*, Heisterkamp et al., 1995), a homologue of the *Drosophila* dishevelled (*dsh*) polarity gene (Pizzuti et al. 1996), and DGCR6, which shares similarity with laminin γ-1 gene (Demczuk et al. 1996)

Aneuploidy: Shapiro and Epstein controversy

The question of how extra or missing copies of a part of the genome (aneuploidy) may cause congenital defects has been the subject of some controversy (Opitz and Gilbert-Barness, 1990). Down Syndrome, the duplication of chromosome 21, has often been used as a model. Unlike syndromes caused by mutations in a single gene, the definite correlation of one or more specific features to an overexpressed gene in an aneuploidy syndrome such as Down syndrome has not yet been successful, primarily because of the large critical region, 400-3000 kb (Rahmani et al. 1990), and therefore, the large number of genes involved. This has inspired the idea that a direct genotype-phenotype correlation in these conditions may not hold true. Shapiro has postulated that the extra copies of the genome in Down syndrome as well as other aneuploidy conditions renders a nonspecific effect on developmental fields, disrupting homeostasis and therefore causing the abnormal phenotypes. One of the arguments presented in support of this hypothesis is the phenotypic overlap between different aneuploidy conditions such as trisomy 21, 13, and 18 (Shapiro 1989). Shapiro proposes that this overlap is generated by the developmental imbalance resulting from the presence of an extra copy of a nonspecific region of the genome.

Figure 1-3: A transcript map of the CATCH22 critical region. The shortest region of overlap (SRO) is the smallest region deleted in all patients. The diagram shows the breakpoint location of the balanced translocation ADU. A second translocation patient, TOH, described in this thesis and in Holmes and Riazi et al. (1997) and the proximal deletion breakpoint of patient G who has mild features of CATCH22, have also been marked in this figure. DGCR3, indicated with a solid box, is located within the DGCR5 gene shown by an open box. This diagram is modified from Dallapiccola et al. (1996), and Gong et al. (1996). The distance between genes and markers is not drawn to scale.

Transcript map of the CATCH22 critical region


~250 kb SRO
CEN
TEL
D22S75
TUPLE1
COMT
LZTR-1
ADU
t(2;22)
TOH
t(21;22)
DGCR6
DGCR5/3
LAN
CTP
CLTCL
DVL22
G patient proximal
deletion breakpoint

On the other hand, there is another school of thought supporting a reductionism view of aneuploidy. This group believes that a direct correlation between the abnormal phenotype and the overexpression of genes in the chromosomal region duplicated in these conditions is indeed possible (Epstein, 1990). Although there are overlapping phenotypes in different aneuploidy conditions, this could be attributed to the fact that many genes are expressed during the development of a particular organ and disruption or change in the expression of any one in a pathway could cause the same abnormal outcome.

Recent findings suggest that at least some of the features seen in aneuploidy conditions could be attributed to one or a few specific genes. Transgenic experiments have allowed the insertion and subsequent expression of a large segment of human DNA in mice (Lamb and Gearhardt, 1995). One such study was done with a YAC fragment containing approximately 180 kb of human DNA from the region of chromosome 21 carrying the gene "minibrain" and little else (Smith et al., 1997). In *Drosophila melanogaster*, the minibrain gene (*mnb*) encodes a serine-threonine protein kinase which appears to play an essential role during postembryonic neurogenesis (Tejedor et al., 1995). The human minibrain gene (MNB) was isolated by positional cloning approaches from the Down syndrome critical region (Shindoh et al., 1996). The transgenic mice carrying this YAC fragment, and hence extra copies of the MNB gene present in the insert, exhibited learning and memory deficits, while the YACs containing unrelated inserts had no effect. This implicates a potential role for MNB in the behavioral phenotype of Down syndrome patients.

Studies of the duplication of 17p associated with Charcot-Marie-Tooth syndrome type 1A have also suggested that the overexpression of the peripheral myelin protein 22 gene (PMP22) in the commonly duplicated 1.5 Mb in 17p11.2-p12 is the underlying cause for the demyelinating neuropathy seen in these patients (Timmerman et al., 1992). Deletion of this gene causes a phenotype known as hereditary neuropathy with liability to pressure palsy (Chance et al., 1993). It seems certain that many specific phenotypes in aneuploidy syndromes should be directly related to over/under expression of dosage-sensitive genes, although general developmental defects and mental retardation may be more difficult to study in this way. Therefore, isolation and characterization of genes in the critical region of syndromes such as CES and CATCH22 will help in understanding of the production of the abnormal phenotype in the aneuploidy syndromes.

Defining the minimal critical region of cat eye syndrome

The critical region for CES (CESCR) is the smallest region of the genome which causes the features of CES when duplicated. Although most of the CES patients carry a

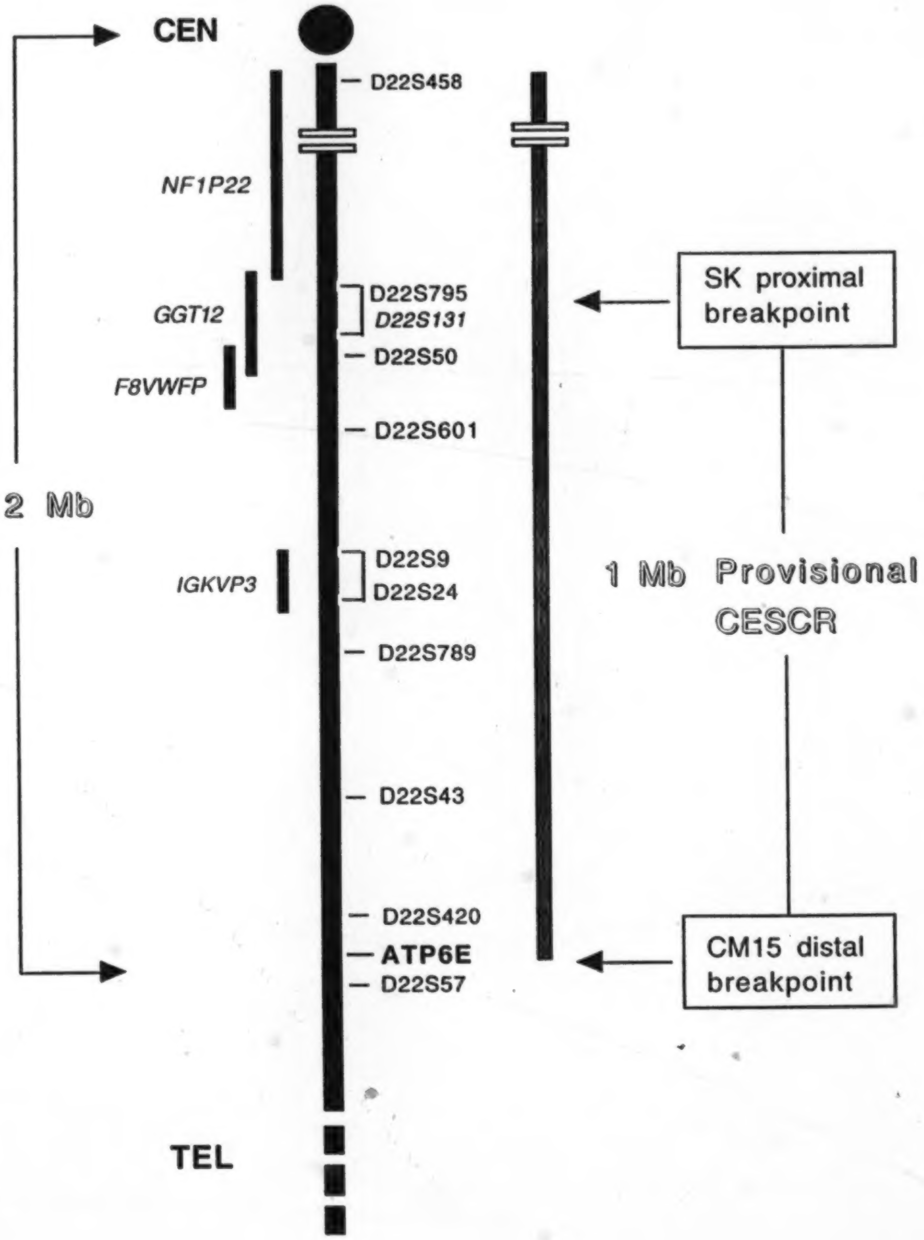
CEC, there are patients with features of CES and having other forms of duplications. These unusual cases may allow the refining of the critical region. One such patient is CM15 (Mears et al., 1995; Mears AJ, PhD thesis, 1995) who carries a supernumerary ring (22). The patient was diagnosed with CES showing coloboma, ear pits and tags, imperforate anus, total anomalous pulmonary venus return, interrupted aortic arch, ventricular septal defect, polycystic kidneys, and urethral reflux (El-Shanti, 1993). The patient died at 17 days, therefore mental development could not be assessed. Using dosage analysis and fluorescence *in situ* hybridization (FISH) it was demonstrated that the breakpoint of the ring (22) was between loci ATP6E and D22S57 (figure 1-4). This delineated the critical region for essentially all the features of this syndrome to a region of approximately 2 Mb from the centromere to D22S57 (McDermid et al., 1996).

Further delineation of the proximal boundary of the critical region was performed using similar studies of patient SK or CE02 (Knoll et al., 1995; Mears AJ, PhD thesis, 1995). Chromosome banding studies showed that this patient carried a visible interstitial duplication of the proximal region of chromosome 22 (Knoll et al., 1995). The abnormal features seen in this patient did not include coloboma and imperforate anus, but did include preauricular pits, total anomalous pulmonary venous return, absent right kidney, absent right testicle, congenital hearing loss, and moderate motor delay. The observation of the features of cat eye syndrome in the patient SK and a second interstitial duplication (LW; Reiss et al., 1985) suggest that three copies of this region are enough to produce the abnormal phenotype associated with CES. Dosage studies demonstrated that the proximal boundary of the duplication was distal to the locus D22S795 (Mears AJ, PhD thesis, 1995), while the distal breakpoint was within 22q12 cytogenetic band. Since D22S795 maps ~1 Mb from the centromere (figure 1-4), this study excluded the region between the centromere and D22S795 from the critical region at least for some features of CES.

The findings with patients SK and CM15 have helped in provisionally mapping the minimal critical region for many features of CES, except possibly for imperforate anus and coloboma, to the ~1 Mb region between D22S795 and D22S57 (figure 1-4). One explanation for the absence of imperforate anus and coloboma in SK may be mapping of these features to the region above D22S795, the region which is not duplicated in this patient. On the other hand this could be explained by the large variation of the abnormal features seen in the CES patients. Also, the duplication in SK includes a large region distal to the CESCR, therefore the effects of the overexpression of the genes located there can not be excluded from having a possible effect on SK. The solution to this problem awaits the

Figure 1-4: Cat eye syndrome critical region (CESCR). The critical region for cat eye syndrome maps to the most proximal region of 22q11, a region of approximately 2 Mb. D22S57 marks the distal breakpoint since this locus is not duplicated in CES patient CM15, while the locus just above it, ATP6E is present in four copies in this patient. The proximal breakpoint for all the features of CES, except possibly anal atresia and coloboma provisionally localizes distal to D22S795, since this locus is not duplicated in SK. ATP6E is the only published gene known in this region. This map is drawn roughly to scale according to McDermid et al. (1996). The four known pseudogenes are shown on the left side of the map and their approximate locations are demonstrated by the solid bars. The tentative localization of these pseudogenes was based on McDermid et al. (1996), Collins et al. (1997), and Hulsebos et al. (1996). D22S131 is present in multiple copies in other regions of chromosome 22.

CAT EYE SYNDROME CRITICAL REGION


CEN
D22S458
NF1P22
SK proximal breakpoint
D22S795
D22S131
GGT12
D22S50
F8VWFP
D22S601
2 Mb
D22S9
D22S24
IGKVP3
1 Mb
Provisional CESCR
D22S789
D22S43
D22S420
ATP6E
D22S57
CM15 distal breakpoint
TEL

study of more patients like SK, who carry an interstitial duplication of this region. For the purpose of this study the critical region was confined to the 1 Mb region between D22S795 and D22S57 (figure 1-4).

The presence of the extra copies of the proximal 22q11 in CES patients suggests the existence of dosage sensitive gene(s) in the region. Although the p arm of chromosome 22 is also duplicated in CES, the p arm of chromosome 22 as well as the other acrocentric chromosomes are thought to contain only repeats and the rRNA genes (Schmickle et al., 1977). In addition the CES features in the two patients with interstitial duplications of the proximal q arm rules out the possibility of the involvement of the p arm in the CES phenotype.

The only gene reported in the provisional ~1 Mb CESCR (figure 1-4) to date is the E subunit of vacular H^+-ATPase (Baud et al. 1994) which is localized proximal to D22S57, approximately 2 Mb from the centromere (McDermid et al. 1996). In addition to ATP6E, a copy of the gamma-glutamyl transferase (GGT) gene family, GGT12, maps to the more proximal region of the CESCR (figure 1-4) (Collins et al., 1995). This is probably a pseudogene, however, since it does not appear to be expressed in the eighteen tissues which were examined by Courtay et al. (1994).

Human genome: gene frequency, distribution, and organization

The human genome, with a size of 3.3×10^9 bp, consists of unique as well as moderately and highly repeated DNA sequences (Lewin, 1996). The sequences are arranged in isochores, long stretches (> 200 kb) of DNA which are compositionally homogeneous (Bernardi, 1985; Saccone et al., 1996). Genomic sequences can be partitioned into five families according to their GC contents: two GC-poor families, L1, and L2, and three GC-rich families, H1, H2 and H3, with H3 being the most GC-rich. L1 and L2 represent about 62% of the genome while H1, H2, and H3 are present in a frequency of 22%, 9%, and 3-4% respectively. The remaining 3-4% is composed of satellite and ribosomal DNAs. By Southern hybridization of probes derived from different genes to different isochores, fractionated based on their buoyant density, it was shown that gene concentration parallels GC levels, being low in GC-poor isochores and increasingly higher in increasingly GC-rich isochores (Bernardi, 1995; Zoubak et al., 1996). The recent analysis of the genomic sequences from the subtelomeric region of chromosome 16 (16p13.3), and chromosome Xq28 region has confirmed the association of genes and the GC-rich DNA sequences (Flint et al., 1997; Chen et al., 1996).

CpG islands are found in the coding sequences of the GC-rich fraction of the genome (Bernardi, 1985). CpG islands, sometimes called HTF islands, were originally identified as short stretches of DNA which contained many sites for the restriction enzyme, *Hpa*II (-CCGG-). These are usually 1-2 kb stretches of DNA with a high proportion of unmethylated GC (~60-70%). CpG islands have been found at the 5' end of all housekeeping genes and many of the tissue-specific genes (Bernardi 1985, Cross and Bird 1995)

The number of genes in the human genome has been estimated to be approximately 50,000-100,000 on the basis of a variety of techniques (Antequera and Bird, 1994; Fields et al., 1994). Out of the approximately $3x10^9$ bp of genomic sequence, only <5% is suggested to code for proteins (Gardiner, 1995). The protein coding fraction of the genome consists mostly of the unique as well as some low copy number repeats which include the gene families (Bernardi, 1995). The genes (at least the ones marked by CpG islands) are not equally distributed on the human chromosomes; rather genes are more abundant in some chromosomes such as chromosomes 19 and 22 (Cross and Bird, 1995; Saccone et al., 1996). In addition the genes are not equally distributed along a chromosome. For example, the telomeric regions of chromosomes are generally believed to be more gene-rich (Saccone et al., 1996; Flint et al., 1997)

The organization of the genomic sequences into cytogenetic bands has been of particular interest. Human metaphase chromosomes can be resolved by different experimental approaches into three main classes of bands, centromeric (C), Giemsa (G), and reverse (R) bands (reviewed in Saccone et al., 1993). Cuny et al. (1981), showed that the R bands are early replicating and GC-rich, thus gene-rich, while G bands are late replicating and gene-poor. In addition to genes, most translocation breakpoints and fragile sites also localize to R bands (Bernardi, 1985). Further studies have demonstrated that approximately 65% of the mapped genes are located in T bands (Holmquist, 1992; Saccone et al., 1993). These are a subset of R bands which are most resistant to heat denaturation and are mainly localized at telomeres (Dutrillaux et al., 1973; Saccone et al., 1996).

Recent advances in large-scale sequencing of the human genome has improved our understanding of gene organization (Flint et al., 1997). Characterization of the genes in 16p13.3 and Xq28, as well as CATCH22 critical region has ruled out the concept of genes being organized into discrete structural and functional units. For example, it was shown that the α-globin regulatory element (HS-40) lies in the intron of a widely expressed gene on 16p13.3 (Flint et al., 1997), and the gene DGCR3 is embedded within the final intron of another gene, DGCR5 (Sutherland et al., 1996). Another finding which has been revealed through the analysis of the genomic sequences is the rare existence of two genes with the same apparent initiation site for transcription on opposite strands. This has been observed for the adrenoleukodystrophy gene and a neighboring gene, as well as for genes *DNL1L* and *XAP-2* in Xq28 (Mosser et al., 1994; Chen et al., 1996)

22q11: gene-rich or gene-poor?

Evidence is accumulating that the long arm of chromosome 22 (particularly q11, and q13) is gene-rich. Using the H3 isochore fraction of human placental DNA for FISH, Saccone et al. (1996) demonstrated that chromosome 22 and 19 are the highest in GC content and therefore very gene-rich, while chromosome 13 and 18 were poorest in GC, and thus relatively gene-poor. Similarly, chromosomes 19 and 22 show very small number of dark G-bands while the gene-poor chromosomes 13 and 18 contain high frequency of G-bands. This parallels the relatively few number of genes assigned to chromosome 13 and 18 (Cross and Bird, 1995), and also is compatible with the live birth of trisomies 13 and 18, while most of the other trisomies usually die during fetal development (McKusick, 1991).

Chromosome 22 is only 53 Mb or approximately 1.9% of the genome (Buetow et al., 1994). In spite of this small size, the number of genes assigned to chromosome 22 to date is 129 (Genome Data Base, June 1997). This represents 5.5% of identified genes, indicating the potentially high number of expressed sequences on this chromosome. With respect to CES, 22q11.2 is an R band and therefore, probably, is gene-rich (Cross and Bird 1995). The establishment of a fairly dense transcript map within the 250 kb CATCH22 critical region surrounding D22S75 (figure 1-3), supports the hypothesis that this region of chromosome 22 is gene-rich with an average of one gene in every 20-25 kb (Gong et al. 1996). Considering the same density of genes for the provisional CESCR, we expect approximately 40 genes in this region, although, the proximity of the CESCR to the centromere may result in a smaller number of genes compared to the CATCH22 region.

Identification of genes by positional cloning

The term "positional cloning" was first introduced by Collins (reviewed in Collins 1995) to describe the identification of genes when location but no functional information about a gene is available. Linkage analysis with polymorphic markers of multiple affected families is usually the first step. Once the gene is confined to a region of an appropriate size (in the range of a few Mb), the gene(s) can be identified using different approaches (below). Finally the association of the gene with the disease can be investigated by sequencing and searching for mutations in the candidate gene. The identification of the cystic fibrosis gene (CFTR) is an example of how positional information was used to isolate genes related to human diseases (Rommens et al., 1989).

Another approach to identify the genes associated with diseases is the "candidate gene approach" where little or no positional information is used. Based on a comparison of the biochemical function of a gene and the phenotype associated with a disease, the possibility of the involvement of a gene is investigated. One example is the demonstration that the dominantly inherited familial cancer syndrome of Li and Fraumeni is due to germline missense mutations in the p53 gene (Sirvastava et al., 1990).

Recently the extensive sequencing of the human genomic and large number of random cDNA clones (ESTs, expressed sequence tags) has allowed the use of this information for identification of syndrome-associated genes. Once a syndrome gene is localized to a region by standard linkage analysis or by other means such as a cytogenetic abnormality, the gene content of the candidate region is investigated for the presence of potential candidate coding sequences. An example of the use of this approach, the positional candidate approach, was the mapping of Marfan syndrome to chromosome 15q and the subsequent identification of mutations in the fibrillin gene, which also maps to 15q, in Marfan patients (Dietz et al., 1991). This approach greatly benefits from the attempts to produce a dense transcript map of the genome. It is believed that by 1998 approximately 80% of the human genes will be mapped and as a result this approach for gene identification will probably overtake the rest (Collins, 1995).

The goal of positional cloning strategies is to identify genes of biological relevance. Positional cloning has already identified the sequences of more than 70 genes associated with specific diseases (http://www.ncbi.nlm.nih.gov/XREFdb/).

Genetic mapping, physical mapping, and cloning of chromosome 22

Botstein et al. (1980) introduced the principles of genetic mapping in human using polymorphic DNA markers. The pattern of inheritance of polymorphic markers can be followed through generations and used to estimate the distance between markers by meiotic recombination frequencies, similar to linkage analysis of different traits in classical genetic analysis. RFLP (restriction fragment length polymorphism) markers were primarily used for this purpose (Lander and Botstein, 1986; Rouleau et al., 1989). These mainly two-allele systems have now been replaced by polymerase chain reaction (PCR)-based microsatellite markers. Genetic mapping of chromosome 22 has produced an estimated size of 98 cM for this chromosome (Buetow et al., 1994). Although genetic mapping has been an invaluable approach for ordering probes along the chromosome, it has some limitations, such as the necessity for the markers to show a high percentage of heterozygosity in order to be informative. Also this technique shows relatively low resolution simply because it depends on the recombination events in meiosis. To achieve the final goal of the "Human Genome Project", which is the cloning and sequencing of the whole human genome, a denser and more accurate map is required. To this end the genetically mapped markers have been used as landmarks for further mapping of STS (sequenced tagged site) markers. An STS is a fragment of DNA with known sequence and for which specific PCR primers have been designed, and is not necessarily polymorphic. In addition to genetic mapping, other means can be utilized to map the anonymous markers including radiation hybrid mapping, somatic cell hybrid mapping, and FISH on metaphase or interphase chromosomes and yeast artificial chromosome (YAC) libraries (reviewed in Bell et al., 1995). Budarf et al. (1996) produced a somatic cell hybrid panel of chromosome 22 using somatic cell hybrid cell lines developed from cells of the patients with chromosome 22 rearrangements as well as cell lines with induced rearrangements of chromosome 22. The hybrid panel divides chromosome 22 into 24 regions (bins) and over 300 markers were then assigned to these intervals. This information was used to integrate the physical and genetic maps of chromosome 22. The STS markers used in this study were also utilized towards assembling a partial YAC contig of chromosome 22 (Bell et al., 1995). YACs can carry large DNA fragments (> 1 Mb), and can be used as a means of ordering markers by producing a contig of a region (Bell et al., 1995; Collins et al., 1995). The pYAC4 vector has been the principal vector used for YAC library construction (Burke et al., 1987). The capability of this vector to carry large inserts as single-copy linear chromosomes in yeast is the result of the centromeric (CEN4) and telomeric (TEL) elements which have been engineered into this vector (Kuhn and Ludwig, 1994). The most updated YAC contig for chromosome 22 has been produced by Collins et al. (1995). They used 620 markers distributed on 22q to isolate YACs by hybridization or screening using PCR. They isolated 705 YACs and assembled them into a high-resolution YAC contig at an average marker density of one per 67 kb.

Despite their advantage in cloning large inserts, the YAC clones suffer from high rates of chimerism, rearrangements, and deletions and are therefore not suitable for sequencing (Hudson et al., 1995; McDermid et al., 1996). In order to circumvent the problems associated with using YACs, recently other types of vectors have been introduced and used in cloning the genomes of human and other model organisms. Currently the YAC contigs are being replaced with more reliable BAC and PAC clones. The BAC (bacterial artificial chromosome) vector is an F factor-based plasmid. It stays in low copy number (one or two copies per cell), thus reducing the potential for recombination between DNA fragments which produces rearrangements within the insert. It is capable of maintaining human genomic DNA fragments of 100-300 kb (Shizuya et al., 1992). In addition to BACs, PAC (P1 artificial chromosome) clones have also been widely used to produce overlapping contigs of the regions of the genome (Kim et al., 1996). These clones replicate using the P1 bacteriophage replication system and can carry inserts larger than 100 kb (Ioannou et al., 1994). Kim et al. (1996) have produced a BAC-based framework contig map of human chromosome 22. These contigs cover approximately 80% of the q-arm of the chromosome. Collins et al. (personal communication) are replacing their YAC map and filling the remaining gaps with these vectors. The generation of such contigs is the first step toward the identification of genes and transcript mapping of chromosome 22. One important feature of the chromosome 22 physical maps revealed by several groups (Collins et al., 1995; Bell et al., 1995) was the high density of low-copy-number repeats on this chromosome, particularity in 22q11. These repeats consist of previously known gene families on 22q11 such as GGT family and BCR-like genes, as well as a number of anonymous low-copy repeat loci such as D22S131 in CESCR (figure 1-4).

Transcript mapping

Many transcript (gene) maps have been produced for those organisms whose genomic DNA has been completely sequenced including 141 viruses, 51 organelles, two eubacteria, one archeon, and one eukaryote, *Saccharomyces cerevisiae* (reviewed in Schuler et al. 1996, Genome Data Base), as well as the partially sequenced genome of *Caenorhabditis elegans* (Blumenthal and Spieth, 1996). A near complete gene map for the human genome will not be available until the sequence of the total genomic DNA is determined, which is predicted to be completed by 2005 (Schuler et al., 1996). Meanwhile, attempts to produce gene maps for various chromosomal segments are underway using a variety of methods. Most of the efforts are concentrated on regions involved in human diseases (Chen et al., 1996). Chromosome 22 has been of major interest because of its numerous associated diseases and cancer-related chromosomal rearrangements (Hudson et al., 1994) as well as its gene-rich status. Partial transcript maps are being produced especially for the regions associated with syndromes (Gong et al., 1996; Dunham, personal communication). In addition to the analysis of the sequenced or partially sequenced contigs, different methods such as exon amplification (Buckler et al., 1991), and direct selection of cDNAs (Parimoo et al., 1991) are being used in order to gain more information about the potentially expressed sequences in specific regions.

Approaches for identification of genes

A. *CpG island identification and cloning*

CpG islands are short stretches of DNA with high frequency (~70%) of unmethylated CpG dinucleotides, versus the bulk of DNA which contains about 40-60% CpG which is commonly methylated (Cross and Bird, 1995). In a study of 375 genes in the GenBank data base, it was found that virtually all widely expressed ("housekeeping") genes and 40% of tissue-specific genes were associated with CpG islands (Larsen et al., 1992). The relevance of the association of these GC rich regions with the 5' ends of genes is not well known. However, Tazi and Bird (1990) have demonstrated that the chromatin surrounding the CpG islands is generally underacetylated, lacks Histone H1 and has a nucleosome-free region indicating a structural effect of the GC rich regions which might facilitate transcription.

The co-localization of CpG islands with the 5' ends of many genes has allowed the utilization of these segments as a signpost for genes. The clustering of unmethylated GC nucleotides generates sites for rare-cutting restriction enzyme, such as *Not*I, and *Eag*I which are methylation-sensitive. This property of CpG islands has been used to isolate CpG island-associated genes (Cross and Bird, 1995). Allikmet et al. (1994) established a *Not*I-linking library of chromosome 3. The sequencing of the clones adjacent to *Not*I sites demonstrated that approximately 90% of these clones contained transcribed sequences. A similar genomic library of *Not*I-linking clones was produced for chromosome 22q11 and a number of putative gene segments were isolated (McDermid et al., 1989). In addition, a PCR-based approach for the isolation of transcribed sequences adjacent to CpG islands has been described (Valdes et al., 1994). This approach is based on the amplification of the region between the cleaved restriction site within CpG islands and a nearby Alu repeat. Alu repeats are a family of interspersed sequences with 300,000 copies in the haploid genome, averaging one for each 6 kb of DNA (Lewin, 1994). Despite being successfully applied in isolation of a number of genes (reviewed in Cross and Bird, 1995), this approach is limited mostly to the housekeeping genes, which carry CpG islands at their 5' ends. Therefore, the genes which are not marked with CpG islands can not be identified by this method. Another disadvantage of this approach is the high background resulting from the CpG islands which are not associated with any genes (Valdes et al., 1994).

B. *Exon trapping*

This approach was introduced by Buckler et al. (1991) for isolation of coding sequences (exons) from cloned genomic DNA by virtue of selecting for functional 5' and 3' splice sites. A pool of genomic DNA fragments from large insert clones (YACs, PACs, BACs, or cosmids) are inserted into an intron present on a specialized mammalian expression vector (pSPL1, pSPL3, or pSPL3B). After transfection, cytoplasmic mRNA is screened by RT-PCR amplification for the acquisition of an exon from the genomic fragment. The splicing donor and acceptor sites in the vector are provided by the intron of the HIV *tat* gene inserted into the rabbit β-globin gene. As well an SV40 origin of replication has been placed upstream of the cloning site. An SV40 viral promoter allows for both replication of the pSPL clones and transcription of the cloned DNAs with no concerns as to developmental and tissue specificity (Burn et al., 1995). The transient expression from this promoter is performed by transfecting COS-1 cells with the constructs. COS-1 cells are African-green-monkey kidney cells that have been transformed by a replication-defective SV40. The mutant SV40 is integrated into the genome of these cells and produces the viral proteins required for a high level of replication and transcription (Gluzman, 1981) from a plasmid such as pSPL3 which contains a wild type SV40 origin of replication and promoter. Although this technique was relatively effective for isolation of genes, the presence of cryptic splice sites within the *tat* intron sometimes generated falsely spliced exons. The original versions of the plasmid (pSPL1, and pSPL3) have now been replaced with pSPL3B. In this plasmid the cryptic splicing site within the HIV *tat* intron has been removed (Burn et al., 1995).

The exon trapping approach has been successfully applied in the isolation of many genes such as the genes responsible for Huntington disease (Huntington Disease Collaborative Research group, 1993) and neurofibromatosis 2 (Trofatter et al., 1993). This approach has the advantage of being expression-independent. However, the amplified fragments are usually small (< 200 bp) and difficult to use for further characterization. Troffater et al. (1995) applied exon amplification to large pools of clones from a flow-sorted human chromosome 22 cosmid library. This resulted in a library of chromosome 22 exons which are now being sequenced and mapped (Troffater J, personal communication). A diagram summarizing the major steps in exon trapping strategy is provided in "Materials and Methods".

C. *cDNA selection*

cDNA selection was introduced as an alternative to the exon amplification technique for the identification of genes from large insert clones such as YACs (Primoo et al., 1991; Lovett et al., 1991). The basis for this "direct selection" scheme is the hybridization of an entire library of cDNAs or cDNAs directly produced from a tissue(s) (Rommens et al., 1993) to an immobilized genomic clone, removal of the nonspecifically bound cDNAs, and further amplification of the clone-specific cDNAs. This technique has been widely used successfully for isolation of genes from numerous regions of the genome. For example, the gene BRCA1 on chromosome 17q21 was identified using this approach (Miki et al., 1994).

Both exon trapping and direct selection have been widely applied in gene isolation experiments. Direct cDNA selection has the advantage of directly capturing the cDNA, which may circumvent the need for further library screening which is required in exon trapping strategy. Another advantage is the ability to capture the genes consisting of only one exon, and therefore, lacking exon/intron boundaries which can be recognized by transcription machinery in the exon trapping approach. A number of genes with a single exon have been isolated from the genome such as SRY gene which is located on the Y chromosome and is involved in the process of sex determination (Sinclair et al., 1990).

In spite of several advantages, direct cDNA selection suffers from being expression-specific. Therefore the genes not expressed or expressed in low abundance in the libraries or tissues used for selection will not be captured, while the exon trapping is expression-independent and would allow the isolation of the genes which are expressed in low abundance (Troffater et al,, 1995).

D. *"UniGene" mapping*

A major part of the "Human Genome Project" is now direct single pass sequencing of cDNA ends from a large variety of cDNA libraries. These are deposited daily into public databases as ESTs (Schuler et al., 1996). Currently the total number of ESTs in database is approximately 450,000 and increasing daily (dbEST, NCBI World Wide Web page). Consequently some genes are represented by multiple ESTs, which may correspond to different portions of a transcript or various alternatively spliced transcripts. To alleviate this problem, recently a group of overlapping ESTs present in database were identified and grouped into "UniGene" sequences (Schuler et al., 1996). The UniGene sequences are available on the World Wide Web (URL: http//www.ncbi.nlm.nih.gov/SCIENCE96/.

These ESTs were mapped using two radiation hybrid panels: the Genebridge 4 RH panel (Gyapay et al., 1996), and the G3 RH hybrid panel (refered to in Schuler et al., 1996), and one YAC panel (Whitehead Institute). This has been a very useful source for identification of the genes involved in the human diseases, such as in the isolation of the gene involved in multiple endocrine neoplasia-type 1 (*MEN1*, Chandrasekharappa et al., 1997) or within the critical regions of many syndromes (Genome Data Base). However, the Unigene database can not be used as the only source of reference since it is unlikely that it equally represents low abundant transcripts. In addition, not all ESTs are represented in the Unigene database. A number of UniGenes have been assigned to 22q11 and fall within the CATCH22 region as well as the CESCR. These genes are currently being characterized further (McDermid et al., unpublished results; Budarf M, personal communication).

E. *Computer analysis of DNA sequences*

A new resource for identification of genes is becoming available as the human genomic DNA is being sequenced. The most efficient approach to identify genes based on sequence data is through the use of the EST database. ESTs matching with the sequence (> 90% identity) are good candidates for genes, since EST sequences are single pass and of low quality sequence. However not all EST cDNAs are true genes. Some represent chimeric clones, or are the result of internal priming within the 3' untranslated regions (Aaronson et al., 1996). In order to search for matching ESTs the sequences can be searched for similarity to previously characterized nucleic acids, using BLASTN, or protein, using BLASTP (URL: http:www.ncbi,nlm.nih.gov). A number of exon prediction/gene modeling programs have been developed to predict coding sequences from eukaryotic genomes, such as GeneID (Guigo et al., 1990), FGENEH (Solovyev et al. 1994), and GRAIL2 (Xu et al., 1994). These programs analyze large genomic DNA sequences and identify genes based on the coding potential of the DNA by comparing it to the DNA and protein sequences in databases. These types of programs are believed to be more efficient than the ones finding genes based on splicing sites predictions (Burset and Guigo, 1995). Ansari-Lari et al. (1997) applied these programs to a 223 kb gene-rich region of chromosome 12p13. Their results demonstrated approximately 60% success in predicting correct exons by FGENEH, and GRAIL2.

The result of many groups shows that the current gene predicting programs should not be used as the only means of identifying genes and it is generally accepted that a combination of the above mentioned approaches as well as the comparative approach which is explained below are essential in order to have a near-complete transcription map of the human genome. A combination of exon amplification and direct cDNA selection has been used in several gene identification efforts (Heiss et al., 1996; reviewed in Brennan and Hochgeschwender 1995), and demonstrated that different sets of the genes were isolated with these methods. This underscores the necessity to use a combination of these gene identification approaches. Currently the Sanger Center (http://www.sanger.ac.uk/HGP/GeneDisc) uses exon amplification and genomic clone sequencing in their gene discovery efforts. A similar approach has been also applied in this study. Direct cDNA selection will no longer be necessary since a large amount of EST sequences are now available in the database which can be searched for similarity to the genomic sequences.

F. *Comparative mapping*

Haldane (1915) was the first to notice the linkage between albino and pink-eyed dilution in laboratory mouse. Later the linkage between phenotypically similar traits was observed in rats (Castle and Wachter, 1924) and deer mice (Summer, 1922). Haldane attributed the linkage of these loci in different species to ancestral chromosomal segments. This concept has been utilized for the study of the human genome, and further improving the transcript maps. The most extensive comparative map made so far is between human and mouse, simply because of the extent of the information on cDNA and genomic sequences available for these two organisms (Epigg and Nadeau, 1995). Haldane's hypothesis predicts that there are conserved synteny and linkage among the genes of the evolutionary-close organisms. "Conserved synteny" refers to two homologous genes that are syntenic in two or more species, regardless of gene order on each chromosome, while "conserved linkage" refers to conservation of both synteny and gene order of homologous genes between species (Eppig and Nadeau, 1995). Copeland et al. (1993) genetically mapped the known mouse genes and demonstrated the presence of 95 regions with conserved synteny and 101 regions of conserved linkages between human and mouse genome, with an average size of 8.9 cM. An updated conserved synteny and linkage map is available through the NCBI World Wide Web page (URL: http://www.ncbi.nlm.nih.gov). This degree of similarity between the human and mouse genomes has proven to be very useful in the discovery of many genes in mouse and human. Galili et al. (1997) sequenced a cosmid derived from the region of chromosome 16 of mouse which contained genes homologous to five human genes mapping to the CATCH22 critical region on chromosome 22. By comparing the sequence of this cosmid with the sequence from the corresponding region in human, a novel human serine-threonine kinase was identified, *Tsk2*, which was homologous to DGS-G gene in the CATCH22 region.

Although there is a conserved linkage between a portion of mouse chromosome 16 and human chromosome 22q11.2 in the region of CATCH22, this does not necessarily apply to all the coding sequences in this region. For example, the gene *CLTCL* which maps within the CATCH22 critical region does not hybridize to mouse genomic DNA (Budarf M, personal communication) and therefore a similar gene in mouse does not exist or it is so diverged it is not detected by the hybridization methods. The disruption of conserved synteny and linkage in comparing a region of mouse and human chromosomes are possibly caused by the rearrangements which occurred in these organisms after the human and mouse ancestors diverged during evolution (Watkins-Chow et al., 1997)

A comparison between the phenotype caused by mutations in mouse and the abnormalities in human syndromes has also been helpful in identification of genes. One example involves the Waardenberg syndrome type 1 (WS1) in humans, the *Splotch* locus in mice, and the *Pax3* gene. Gene mapping localized *Pax3* to a region of mouse chromosome 1, and making it a candidate for the Splotch mutation (Goulding et al., 1991). When WS1 was mapped to the homologous portion of human genome, 2q37 (Foy et al., 1990), *PAX3* also became a candidate for WS1. Subsequent studies also demonstrated *PAX3* mutations in WS1 (Tassabehji et al., 1992).

A comparative map of the proximal regions of human chromosome 22q11 with mouse chromosomes shows synteny of 22q11 to parts of mouse chromosome 16, 10, and 5 respectively from proximal to distal (Galili et al., 1996; Genome Data Base, figure 1-5). The region of the mouse chromosome with conserved linkage to the CESCR has not yet been identified.

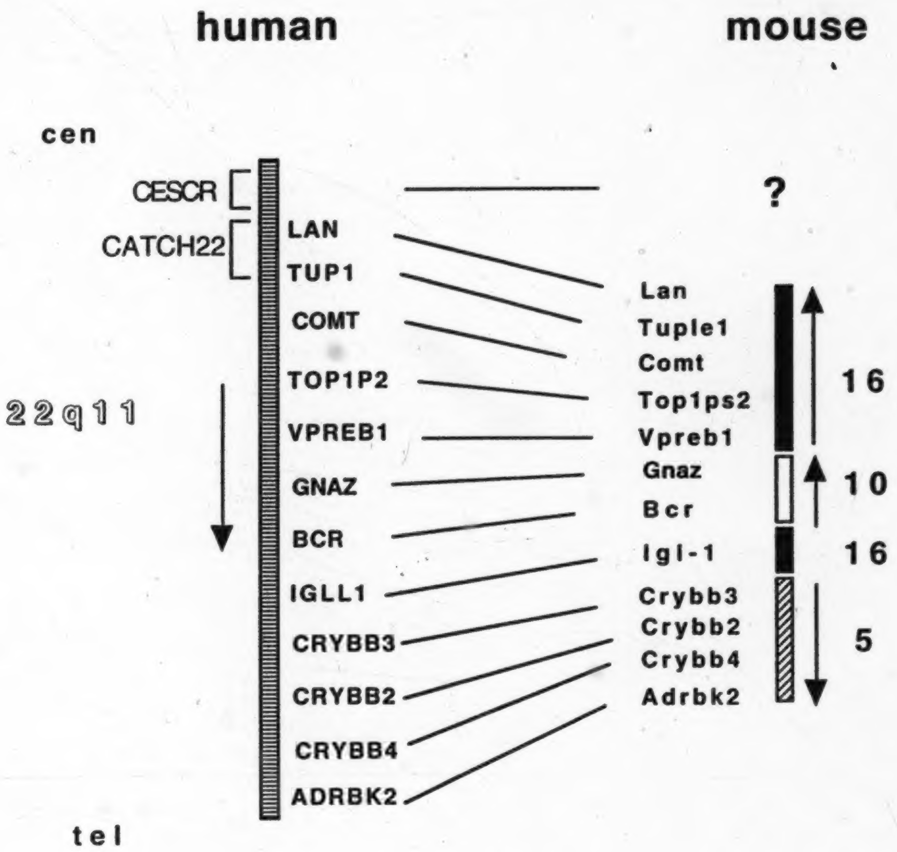

human
mouse
cen
CESCR
CATCH22
22q11
tel
LAN
TUP1
COMT
TOP1P2
VPREB1
GNAZ
BCR
IGLL1
CRYBB3
CRYBB2
CRYBB4
ADRBK2
?
Lan
Tuple1
Comt
Top1ps2
Vpreb1
Gnaz
Bcr
Igl-1
Crybb3
Crybb2
Crybb4
Adrbk2
16
10
16
5

Figure 1-5: A human/mouse comparative map of 22q11. The map spans from the CATCH22 critical region to the locus, ADRBK2. The genes in this region map to three different mouse chromosomes, 16, 10, and 5. The arrows indicate the centromere to telomere direction. The mouse chromosomal region with a conserved linkage to the CESCR has not yet been identified. These data were collected from NCBI World Wide Web page (URL: http://www.ncbi.nlm.nih.gov)

Research objectives

A. The first goal was to help produce a YAC contig of CESCR, to be used as a source for further isolation of more reliable clones such as cosmids. Recently a more useful BAC/PAC contig encompassing the provisional critical region has been made. These resources were used in this study for isolation and characterization of genes.

B. The main purpose of this project was to isolate and characterize genes in the cat eye syndrome critical region. This is a region of approximately 2 Mb at the pericentomeric region of 22q, as identified by patient CM15. The study of the interstitial duplication patient SK further refined the critical region for most of the features of this syndrome to a smaller 1 Mb region, from D22S795 to the locus D22S57. Most of our efforts towards identification of genes have been concentrated on this provisional critical region. Here I present my results for the isolation and characterization of some of these putative genes. To this end, "exon amplification" was applied to some of the clones in CESCR. The isolated exons were characterized by Southern and Northern hybridization, sequencing, RT-PCR, and whole mount mouse *in situ* hybridization. These results are presented in parts B and C in chapter III.

A second objective was to determine the region of the mouse genome with conserved synteny to CESCR. The putative gene segments were used to isolate mouse genomic phage and BAC clones. Later these clones were mapped by FISH to mouse chromosomes. This information will be useful towards identification of new genes from this region by comparative studies.

C. The final goal of this study was to localize the chromosome 22 breakpoints in seven patients with balanced translocations involving 22q11.2. This was done by FISH using cosmids in the region of the breakpoints. This study was performed to investigate clustering of the breakpoints and any possible relation to the rearrangements in CES or CATCH22. We were specifically interested in mapping the translocation breakpoint in the patient TOH who demonstrates some features of the CATCH22 syndrome.

Chapter II: Materials and Methods

Tissue culture

Standard methods of cell culturing were used. All human cell lines were grown in a 37°C humidified incubator with 5% carbon dioxide in T25 tissue culture flasks (Corning). The lymphoblastoid cell lines were maintained in RPMI 1640 media (Gibco/BRL) supplemented with 10% fetal bovine serum (FBS, Gibco/BRL), 1% L-glutamine (200 mM stock, Gibco/BRL) and 1% penicillin/streptomycin (stock 5000 U/ml, Gibco/BRL). When necessary the culture was passaged by dilution of 1:3 in fresh medium. Fibroblasts were cultured in DMEM media (Gibco/BRL) supplemented with 10% FBS, 1% L-glutamine, and 1% penicillin/streptomycin. To passage the fibroblast cultures, the cells were detached from the surface of the flasks by washing first with Hank's Balanced Salt Solution (Gibco/BRL) to remove the trypsin inhibitors, and then by adding a minimal amount of trypsin-EDTA (BRL). The cells were then diluted in supplemented DMEM in several T25 or T75 flasks.

To freeze lymphoblastoid or fibroblast cultures the cells were pelleted by centrifugation at 130 *g* for 10 min. The cells were resuspended in the freezing medium (FBS with 5% dimethyl sulfoxide from Sigma) at a final concentration of 5×10^5 cells/ml, transferred into cryotubes (Nunc.), and placed in the gas phase liquid nitrogen for an hour before they were transferred to a liquid nitrogen tank.

Genomic DNA extraction

For genomic DNA extraction a modification of the procedure described by Gustincich et al. (1991) was used. The original method was modified by Mears et al. (1994) in order to prepare DNA from a larger volume of blood as well as from lymphoblasts and fibroblasts. In this study genomic DNA was extracted from lymphoblastoid cell lines. Three to four T25 flasks of the cells were grown, as described. Cells were pelleted by centrifugation at 800 rpm for 10 min, and resuspended in 5 ml of Hank's balanced solution. Then 10 ml of the lysis buffer was added (8% DTAB (Sigma) in 1.5 M NaCl, 100 mM Tris-HCl pH 8.6, 50 mM EDTA) and the mixture was incubated for 15 min at 65°C. Then 15 ml chloroform was added, the tubes were mixed by inverting 3-4 times and centrifuged at 10,000 rpm for 15 min. The aqueous layer was then transferred to a fresh tube and 17 ml of dH_2O + 1.7 ml of CTAB (Sigma) (5% CTAB, in 0.4 ml NaCl) was added, mixed and spun at 1000 rpm for 10 min. The pellet was resuspended in 5 ml of

1.2 M NaCl for 2-4 hr before adding 18 ml of cold 95% ethanol to the tube. The DNA was removed from the tube using a toothpick, or spun briefly using a clinical centrifuge and the aqueous layer was decanted. DNA was washed with 70% ethanol. Typically the DNA was resuspended in 500 µl of TE (pH 8.0), and a total yield of 100-500 µg of DNA was obtained.

Plasmid and cosmid DNA preparation

Small scale DNA preparations were performed on 5 ml overnight bacterial cultures in LB (Luria-Broth) using the alkaline lysis method as described in Sambrook et al. (1989). For large scale DNA preparation, a commercially available kit (Qiagen-tip 500) was used. Overnight cultures of 100 ml and 500 ml were used respectively for plasmids and cosmids in order to prepare 200-1000 µg of DNA. Cosmids were all isolated from the chromosome 22-specific cosmid library LL22NC03 (de Jong et al., 1989).

Approximately 3 µg of the DNA preparations were digested with different restriction enzymes and electrophoresed in an 0.8% agarose gel to check for the yield and size of the insert. The restriction digestion reactions were performed as per manufacturer's instruction.

Preparation of bacterial artificial chromosome (BAC) DNA

For large-scale DNA preparation of BAC DNA a procedure originally described by Shizuya et al. (1992) was used. Cultures were grown in 100 ml LB containing 12.5 µg/ml chloramphenicol for overnight at 37°C. Cells were pelleted by centrifugation at 5000 rpm for 5 min, resuspended in 10 ml of solution 1 (25 mM Tris.HCl, 50 mM glucose, 10 mM EDTA), with lysozyme added to a final concentration of 2.5 mg/ml. Tubes were transferred to ice and 20 ml of freshly prepared solution 2 (0.2 M NaOH, 1% SDS) was added and incubated for 5 min. Then 15 ml of solution 3 (5 M potassium acetate pH 4.8) were added to the tubes to precipitate the proteins and genomic DNA. The samples were centrifuged for 15 min at 10,000 rpm to pellet the flocculent precipitate, and the supernatant was transferred to another tube. The supernatant was treated with 10 µl of 10 mg/ml of RNase A (Sigma) for 30 min at 37°C. Finally BAC DNA was precipitated by adding 1 volume of isopropanol and centrifugation at 10,000 rpm for 30 min. The DNA pellets were washed with 70% ethanol and dissolved in 500 µl of TE pH 8.0. BAC DNAs were digested with different enzymes and run on a 0.8% agarose gel for further analysis.

Preparation of yeast artificial chromosome (YAC) DNA and PFGE of YACs

A protocol described by Rose et al. (1990) for isolation of yeast DNA was used. YAC clones were cultured in 40 ml of YPD (yeast complete media) at 30°C. Due to instability of YACs and the possibility of deletion within the inserts after long incubation time, the YAC clones were cultured for a minimum incubation time, usually 2 days.

Agarose embedded YAC DNA for pulsed field gel electrophoresis (PFGE) was prepared as described in Rose et al. (1990). The agarose plugs containing YAC DNAs were electrophoresed using a pulsed field gel electrophoresis CHEF-DRII apparatus. The plugs were loaded into a 1% agarose gel and electrophoresed in 0.5x TBE at 14°C at 200 Volts with a switch time of 60 s for 14 hr followed by 90 s for 8 hr. The YAC DNA was transferred to nylon membrane as explained below and probed with α^{32}P-labeled total human DNA to size the YACs.

Southern blotting and hybridization of DNA probes

For Southern blotting of genomic DNAs the gel was denatured, neutralized and then transferred to GeneScreen*plus* nylon membranes (DuPont), according to the manufacturer's instruction. For plasmid or cosmid DNA, an alkaline transfer method was used according to the manufacturer's instruction (DuPont). To make DNA probes, DNA fragments were purified in 0.8% low-gelling temperature agarose (Sea Plaque, FMC), boiled for 10 min and then labeled by random priming (Feinberg and Vogelstein, 1984) using $\alpha[^{32}P]$ dCTP (ICN). Usually 20 µl (50-100 ng) of denatured DNA in agarose was added to 10 µl of oligo labeling buffer (100 µl 1 M Tris, 12.5 µl 1 M $MgCl_2$, 2.5 µl 50 mM dATP, 2.5 µl of 50 mM dGTP, 2.5 µl of 50 mM dTTP, 250 µl 2 M Hepes pH 6.6, 150 µl 90 A260 units/ml random primers, and 1 µl of β-mercaptoethanol); 2 µl of 10x BSA (BRL); 5-7 µl of $\alpha[^{32}P]$ dCTP (50-100 µCi); sterile water to 48 µl; and 2 µl (20 units) of Klenow fragment (Pharmacia), followed by incubation for 2-4 hr at 37°C. The reaction was passed through G-50 Sephadex to remove the unincorporated nucleotides. Probes containing repetitive sequences, were boiled for 10 min with 500 µl of sonicated human placental DNA (2.5 mg/ml) and 84 µl of 1 M sodium phosphate (pH 7.0) and incubated for 1 hr at 65°C to preanneal the repetitive portion of the probe (Litt and White, 1985). The probe was then added to the Southern blots in a rolling bottle hybridization oven (Tyler) in the hybridization solution (10% SDS, 1 M EDTA, 1 M $NaPO_4$, 5x Denhardt's). The hybridization and washing conditions were as per the manufacturer's instruction. For "Zoo blots", when cross hybridization of human probes to DNA from other organisms or vice versa was desired, a lower temperature of hybridization and washing, 50-55°C, was used.

Human and mouse cDNA and genomic library screening

A list of the cDNA libraries as well as a BAC library of mouse genomic DNA, and their sources used in this study, has been presented in table 2-1.

Plating the cDNA libraries

The bacterial plating cells used were either XL1-Blue (*recA1, endA1, gyrA96, thi-1, hsdR17, supE44, relA1, lac,* [F' *proAB, lacI*qZDM15, Tn*10 (tet^r)*], Stratagene) for λZapII clones or LE392 (*supE44, supF58, lacY1 or Δ(lacIZY)6, galK2, galT22, met B1, trpR55, hsdR15 F-*) for λgt10 and λgt11 clones. The libraries were titered by plating different dilutions on LB plates. Plating the phage clones was done by standard procedure (Sambrook et al., 1989). A total of approximately 1x10^6 phages from different libraries were plated on LB plates (150 x 15 mm size), at an approximate density of 5x 10^4 plaques per plate. For the plasmid kidney library between 10^4-2x 10^4 colonies were plated on LB plates containing ampicilin (50 μg/ml), and tetracycline (10 μg/ml).

Plaque and colony lifting and hybridization

The plates were chilled at 4°C for 30 min. Plaque or colony lifting was performed using Hybond^{Tm} nylon membranes according to the manufacturer's instruction (Amersham Life Science). Membranes were denatured in 1.5 M NaCl/0.5 M NaOH, neutralized in 3 M NaCl/0.5 M Tris-HCl pH 8.0, and baked at 80°C for 2 hr to fix the DNA to membrane. The hybridization solution consisted of 5x SSC, 5x Denhardt's solution, and 0.5% (w/v) SDS. The probe was added to a final concentration of approximately 1x 10^6 cpm/ml and hybridization was performed at 65°C, or 55°C when screening a mouse cDNA library with a human probe. The primary positives were plated at low density and rescreened with the same probes to confirm the result. A third round of plating and screening was sometimes necessary to isolate the phage or bacterial clone positive for the probe.

Table 2-1: Genomic, and cDNA libraries used in this study.

VECTOR	TISSUE	SOURCE
BAC	mouse genomic DNA	Dr. B. Birren
pcDNA1	human adult kidney	Clontech Inc.
λZapII	human fetal brain	Stratagene Inc.
λgt11	human fetal liver	Clontech Inc.
λgt11	human adult heart	Dr. R. Farahani
λZapII	mouse brain	Dr. J. Rommens
λgt11	mouse fetus, 8.5 days	Dr. D. Rancourt
λgt11	mouse fetus, 10.5 days	Dr. D. Rancourt
λgt10	mouse fetus, 12.5 days	Dr. D. Rancourt
λZap	human Caco-2* cell line	Dr. J. Rommens

* Caco-2 cDNA library is derived from a colon carcinoma cell line.

Screening a mouse BAC genomic library

A BAC library of mouse genomic DNA on high-density membranes was kindly provided by Dr. B. Birren (Division of Biology, California Institute of Technology). The library was screened with a human probe by hybridization at 50°C to allow cross-hybridization of the probe to the mouse genome. The hybridization and washing solutions were essentially the same as was used in Southern hybridization. The hybridization was performed in 23x 23 cm containers in a shaking incubator (Innova 4080, New Brunswick Scientific).

Preparation of total and poly A$^+$ RNA

For this purpose RNase-free conditions were used. To remove any RNase contamination glassware was filled with 0.2% DEPC (diethyl pyrocarbonate, Sigma) treated water, left at room temperature (RT) for at least 2 hr and then autoclaved. Plasticware (polypropylene and polyethylene tubes and transfer pipets) was washed with chloroform and then rinsed with RNase-free water. To prepare RNase-free water, 2 ml of DEPC was added to each liter of water and left at RT for at least 2 hr before it was autoclaved. RNA was extracted from human and mouse tissues as well as cell lines. The human adult or fetal tissues were obtained from the Departments of Pathology or Pediatrics, University of Alberta, and kept in liquid nitrogen before they were transferred to a -70°C freezer for longer storage. Mouse tissues were transferred to dry ice upon removal and stored at -70°C until needed. Approximately 1-5 g of tissues were homogenized in a 10x volume of TriZol solution (BRL) using a homogenizer in an RNase-free 50 ml tube (Corning). For cell lines, 1 ml of TriZol was used for each 5-10x 10^6 cells. The RNA extraction was performed as per the manufacturer's instruction. The RNA was usually dissolved in 500 μl of RNase-free water. The total yield varied for different tissues with an average yield of 500 μg of RNA from 1 g of tissue. The "PolyATtract mRNA Isolation System IV" purchased from Promega Inc. was used to isolate messenger RNA (mRNA) from 0.5-1 mg of total RNA. The poly A$^+$ RNA was hybridized to biotinylated oligo(dT) and captured using streptavidin-coupled paramagnetic beads. The extracted total RNA was run on an agarose gel in order to check for the integrity of the RNA.

Northern blotting

From 2-10 µg of the extracted mRNA or total RNA respectively was electrophoresed in 1.2% agarose gel in RNase-free 1x MOPS at 5 V/cm according to Ausubel et al. (1989). An RNA ladder (BRL) was included and stained with Ethidium Bromide to measure the size of the RNAs. The RNA was transferred to nylon membranes and blotted in 10x SSC. Alternatively human adult and fetal multiple tissue Northern blots were purchased from Clontech Inc. Human exons or cDNA fragments were labeled and hybridized to the membranes at 42°C overnight. The hybridization solution was 50% formamide, 5x SSPE, 10x Denhardt's solution, 2% (w/v) SDS, and 100 µg/ml salmon sperm DNA. The membranes were washed as follows: 2x 5 min in 2x SSC, 0.1% SDS at RT, and 15-30 min in 0.1x SSC, 0.1% SDS, all at 42°C. When cross-hybridization between human and mouse was desired, hybridization and washing were performed at 37°C. Blots were exposed to Kodak X-OMAT film at -70°C for the appropriate time.

Polymerase chain reaction

Polymerase chain reaction (PCR) was performed using a PTC-100™ thermocycler (MJ Research Inc.). Typically a 50 µl reaction was set up in 1x buffer (50 mM KCl, 10 mM Tris.HCl, 1.5 mM $MgCl_2.6H_2O$), 0.2 mM dNTPs), and using 10-20 pmoles of each primer (table 2-2). Primers were either made at the Oligonucleotide Synthesis Facility at the Department of Biological Sciences, University of Alberta, or were ordered from Bioserve Inc. The PCR conditions (denaturing, annealing, extension, and the number of cycles) were varied according to the GC content of primers and the size of the product. For PCR amplification using tyrosine kinase degenerate primers, the reactions were denatured for 4 min at 94°C, followed by 30 sec. denaturation at 94°C, 1 min annealing at 55°C, and 1 min extension at 72°C, for a total of 35 cycles.

Reverse transcription-PCR (RT-PCR)

Approximately 2 µg of total RNA extracted from a 9 wk old human embryo was subjected to RT-PCR. Twenty pmoles of a gene specific primer and DEPC-treated water was added to 12 µl. The mixture was incubated at 70°C for 5 min and chilled on ice for 1 min. Then 4 µl of buffer (250 mM Tris-HCl (pH 8.3), 375 mM KCl, 15 mM $MgCl_2$), 2 µl

Table 2-2: List of the primers used in this study.

* TKd1 and TKd2 are degenerate primers which have mixed bases at certain positions: A/G=R; A/C=M; A/T=W; G/C=S; G/T=K; C/T=Y; A/G/C/T=N

genes/exon	primer	5'-3' sequence
KCNMBL/KCNMBLP	RE1	AGGTTCACAAACACCCGAAGAC
	LE1	GAGAGAAGAATCAACTGCACTG
	P1	GAGACCCACTAGATGTGCAC
	P2	TGAGCATAAAAGCCTTTAGAATGG
	RE1B	GCACCACCTAGCAGAGTCAG
	RE1B'	CTGACTCTGCTAGGTGGTG
	RE1-E	GCAGACAGGCATAATTAGGTCC
	R1	ACACACCTAAGTGCCACCAAG
SAHL1	11-1	TGGTGAAGGCATTGATTGT
	11-2	ACAATCAATGCCTTCACCA
	22-1	CAGAGAACACAGCCGAAGTG
	22-2	CACTTCGGCTGTGTTCTCTG
	PJ11	CTTCACCACCCAGAGGCATTG
	PJ22	TCTGGGTTACCCGCCGAGGC
IDGFL/e5	B-1	GCAATGTCTTTCTTCGGCTA
	B-2	GACTATAAGGGTGGGTCGTA
e7	7-1	GCAGAAGGCAGCACAGACCAT
	7-2	CTTGATCAGGGCCTGAGATGC
e8	8-1	CAGAAGACAGCGGAGAGAGG
	8-2	ACCGCGTAGTGGATGTTACC
e6	E-1	GCGACGATTGTCTGTTTCTTCTT
	E-2	AAGGAGAAACAGACAATCGTCGC
CES-TK	TKd1	RTYMRRTGGAYSGCCCYKGA
	TKd2	AYGCCAWAGGMCCANACATC
	TK-1	GGACAAGCTTGTTCACCGAGA
	TK-2	CTCTAGCTTGGACCATTCACCG
FB20C7	Fb-1	TTGAATGTGCCAGGCATGAGGC
	Fb-2	GGGAGGTGCAAGACTCTTGGC
3' RACE	AN	GACTCGAGTCGACATCGATGCA
	AN$(T)_{17}$	GACTCGAGTCGACATCGATGCA$(T)_{17}$

of 100 mM DTT, and 1 μl of 10 mM dNTPs were added to the mixture and it was placed at 42°C for 2 min. Finally 1 μl (200 units) of Superscript II Reverse Transcriptase (BRL) was added to the reaction and incubated at 42°C for 30 min. The reverse transcribed RNA was digested by treating with 1 μl (3 units) of RNase H at 55°C. Typically, 5 μl of the reverse transcription reaction was used in a PCR reaction.

Rapid amplification of 3' end of cDNA (3'-RACE)

3'-RACE was performed using a procedure originally described by Frohman et al. (1988). This was done to amplify and clone the 3' end of KCNMBL in the vicinity of D22S795 (figure 2). Approximately 2 μg of human total fetal RNA was reverse transcribed using a primer anchored with a run of 17 Ts ($AN(dT)_{17}$, table 2-2). The reverse transcription and RNase H treatment conditions were as described before. $AN(dT)_{17}$ primers which were not hybridized were removed from the reaction by using a Glassmax column (BRL). Five μl of the purified first strand DNA was amplified for 30 cycles using a gene specific primer (LE1, table 2-2) and $AN(dT)_{17}$. A dilution of the primary PCR reaction was amplified with a gene specific nested primer (LE1', table 2-2) internally located compared to LE1, and the anchor primer with no Ts at the end (AN, table 2-2). The annealing temperature was 60°C so to prevent non-specific binding of the primer. The amplified DNA was run on a gel and the band was eluted from the gel using a Glassmax column (BRL). The DNA was then cloned into pGEM-T vector (Promega) for further analysis. This vector is specifically designed for the subcloning of PCR-amplified DNA fragments, by digesting the $pGEM^R$-5ZF(+) vector with *Eco*RV and adding a 3' terminal thymidine to both ends (Mezei and Storts, 1994).

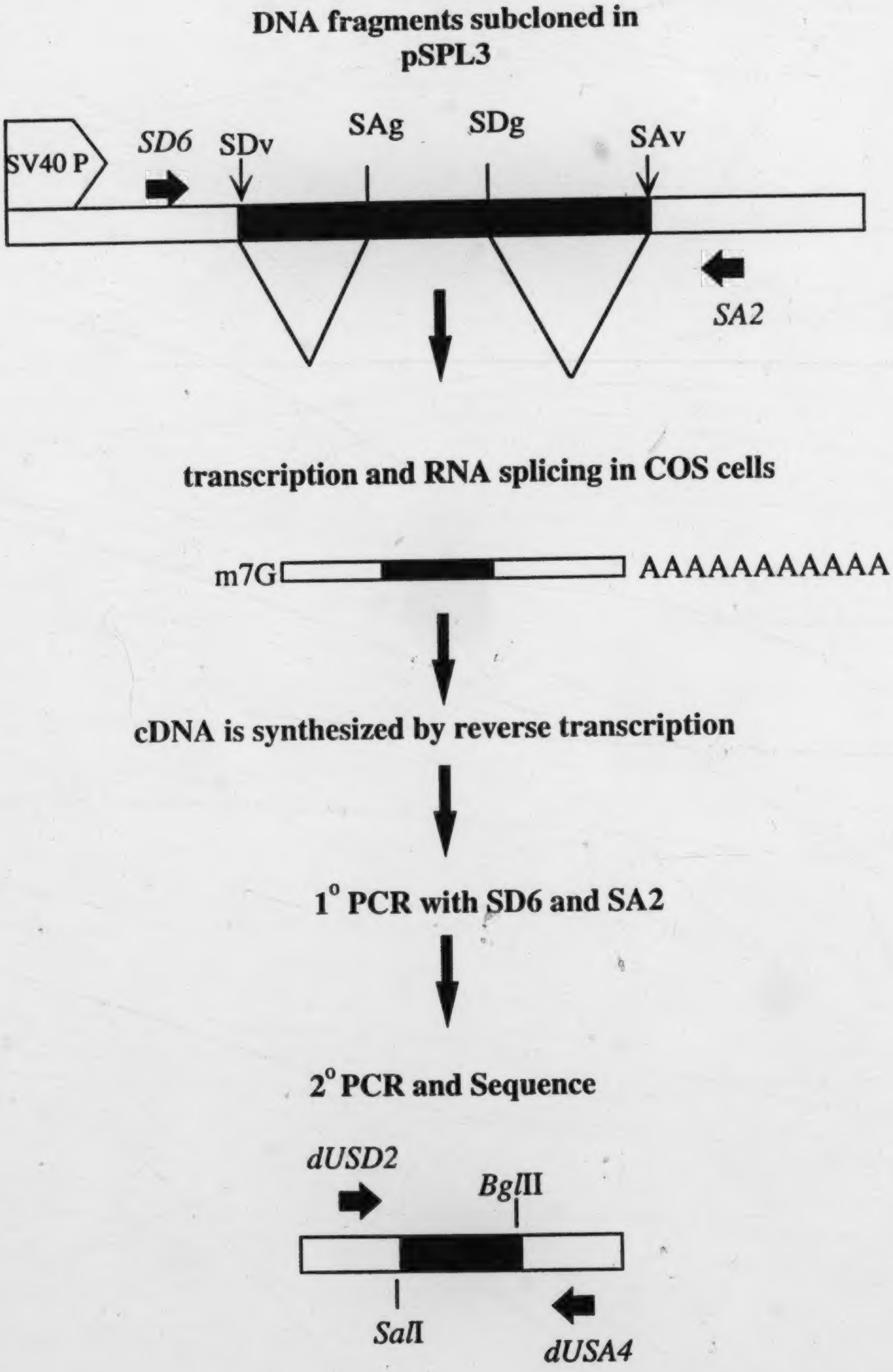
Exon trapping

Exon trapping was performed using a kit (Exon Trapping System, BRL) with some modifications. Figure 2-1 demonstrates an overview of the exon trapping strategy.

1. *Subcloning into pSPL3 or pSPL3B*

The genomic DNAs (cosmids, BACs, and PACs) were digested with either *Sst*I or *Bam*HI restriction enzymes and randomly cloned into pSPL3 and later pSPL3B cloning vector. This vector carries a region of the *tat* gene from HIV which contains splicing donor and acceptor sites, as well as an origin of replication and transcription from SV40 virus.

Figure 2-1: Schematic representation of the exon trapping procedure. The human DNA was randomly subcloned into pSPL3, and transiently expressed in COS-1 cells. The cytoplasmic RNA was extracted and reverse transcribed. Finally the putative exons were amplified in two rounds of PCR. Exons were treated with *Bgl*II and *Sal*I restriction enzymes which release most of the flanking region of the exons, leaving a total of 18 bp of the vector sequence. SDv: splicing donor of vector, SAv: splicing acceptor of vector, SDg: splicing donor of genomic insert, SAg: splicing acceptor of genomic insert.


DNA fragments subcloned in pSPL3
SV40 P
SD6
SDv
SAg
SDg
SAv
SA2
transcription and RNA splicing in COS cells
m7G
AAAAAAAAAAA
cDNA is synthesized by reverse transcription
1° PCR with SD6 and SA2
2° PCR and Sequence
dUSD2
BglII
SalI
dUSA4

The clones were grown overnight and plasmid DNA was prepared from the cultures as explained before. Alternatively, the clones were plated on LB medium. The colonies were transferred to Hybond^{Tm} nylon filters and probed with the α^{32}P-labeled total human genomic DNA in order to isolate the colonies carrying the fragment with human inserts. This was done to remove the clones which contained contaminating fragments from the cosmid or BAC vector, as well to exclude the clones derived from the self splicing of pSPL3 or pSPL3B splicing donor and acceptor sites.

2. *Transformation, RNA extraction, and reverse transcription*

The plasmid DNA prepared from shotgun cloning was electroporated into COS-1 cells using an Electro Cell Manipulator (BTX Inc.). The growth conditions for COS-1 cells are the same as fibroblasts. Approximately 1-5 x 10^6 cells at 60-70% confluency were used for each experiment. The cells were grown in supplemented DMEM for 2-3 days in order to reach the required confluency. The COS-1 cells were harvested using the same methodology as described before for fibroblast cell cultures. Cells were resuspended in 400 ml of DMEM or PBS (8.1 mM $Na_2HPO_4.7H_20$, 1.2 mM KH_2PO_4, 2.7 mM KCl, 138 mM NaCl), checked for concentration using a hemacytometer and transferred into cuvettes (0.4 cm electrode gap, Bio-Rad Laboratories). Approximately 1 µg of DNA was added to the electroporation cuvettes, which were chilled on ice for 5 min. The electroporation was performed at 500 Volts and 1000 µFaradays to yield a time constant of 10-15 msec. Cells were returned to the ice for a few minutes and then transferred to T25 tissue culture flasks containing fresh supplemented DMEM and incubated for overnight. The next day the cells were harvested and subjected to cytoplasmic RNA extraction.

Approximately 1 µg of the extracted cytoplasmic RNA was used for reverse transcription using primer, SA2 (figure 2-1), which hybridizes to a region of the flanking tat gene. The reverse transcription was performed using Superscript II RT (BRL).

3. *Polymerase chain reaction to amplify the putative exons*

The reverse transcribed templates were first amplified with primers SD6 and SA2 (figure 2-1) for six cycles. The reaction was then treated with *Bst*XI enzyme to digest possible PCR products resulting from the splicing of the vector ends. This BstXI site is produced after splicing of the vector's donor and acceptor splicing sites. A second round of amplification (30 cycles) was performed with nested primers dUSA4 and dUSD2 (figure 2-

1) to amplify the exons. Finally the putative exons were run on a 1.5% agarose to separate the fragments.

4. *Confirming the origin of the exons*

In order to confirm that the amplified fragments originated from the clone used, the putative exons were run on an agarose gel and then Southern blotting was performed. The blot was hybridized with the α^{32}P-labeled genomic clone which was originally used for the isolation of the exons. If necessary the exons were hybridized back to the original clone by Southern hybridization. To exclude any amplified DNA which originated from the pSPL3B vector itself, the blot was also hybridized with the 2.4 kb intron of the HIV *tat* gene in the vector. For further characterization of the exons on Southern or Northern blots, the exons were digested by *Bgl*II, and *Sal*I to remove the 177 bp flanking sequence from vector which is amplified after the secondary PCR amplification with primers dUSA4 and dUSD2. This was necessary because this fragment hybridized in my Northern blots strongly to a 0.9 kb transcript of unknown origin in the liver.

Sequencing of the exons and cDNAs

The isolated exons were sequenced primarily by the automated sequencing facility in Dr. Roy's laboratory (Department of Biological Sciences, University of Alberta) using an ABI 373A. Subsequently the exons and cDNAs were either manually sequenced by D. Shkolny and A. Johnson (Department of Biological Sciences) using a cycle sequencing kit (USB) and α^{33}P[dCTP], or they were sequenced in collaboration with Dr. N. Shimizu's laboratory (Department of Molecular Biology, Keio University, Japan)

Database searching

A number of computer programs were used to search for DNA and protein similarities between the sequence produced and known DNA sequences in various databases. For this purpose, the BLAST series of programs provided by the National Center for Biotechnology Information (NCBI) were used. BLAST (Basic Local Alignment Search Tool) is the search algorithm employed by several programs, such as BLASTN (DNA sequence) and BLASTP (amino acid sequence). A number of BLAST programs are available through the Baylor College of Medicine's BCM search launcher home page (URL: http://gc.bcm.tmc.edu:8088/). Typically, a DNA sequence from an exon or cDNA was primarily searched by BLASTN and BLASTN/dbest. The search was done using the default parameters which are set at the expected value of 10 (the statistical significance threshold for reporting matches against database sequences) and an alignment value of 50 (which determines the number of sequences reported based on their high-scoring segment pairs) (BLAST manual, available at NCBI BLAST World Wide Web page). BLASTN searches the similarity to all DNA sequences in the database while BLASTN/dbest looks for similarity to ESTs (expressed sequence tags). If the similarity to the EST sequences in the database (usually >150 bp) was over 95% the ESTs were further characterized. EST clones were purchased from either Genome System Inc. or American Type Culture Collection (ATCC). BLASTX, TBLASTX, or BEAUTY (Blast enhanced alignment utility, Worley et al., 1995) searches were performed to identify protein homology and the presence of any known functional domain.

The genomic sequence of cosmid c1A3, which is located in the vicinity of D22S9 and contains part of the gene SAHL1 identified in this study (figure 3-11) was further analyzed for localization of potential coding sequences. Primarily the Grail-2 program was used to predict exons (BCM search launcher). As well, the cosmid sequence was submitted to the Genome Annotation and Information Analysis server (GAIA, http://agave.humgen.upenn.edu:80/gaia/) for further analysis and mapping of the ESTs and other landmarks on this cosmid sequence. GAIA is a data analysis and storage system for genomic sequence and its annotation, which is being developed at the Center for Bioinformatics, University of Pennsylvania.

Mouse whole mount *in situ* hybridization

The protocol described in Wilkinson (1987) was used with some modifications.

1. *Collection and storage of embryos*

We established a small colony of mice, CD1 strain (Charles River Canada). This is an inexpensive outbred strain with a large litter size. When required, the mice were mated overnight and the female mice were examined the next morning for the presence of a mating plug. If the plug was observed, the day was considered 0.5 days post conception (0.5 dpc). Mice were euthenized and embryos were removed at 7.5, 9.5, 11.5, 12.5, and 18.5 dpc.

The embryos were collected in cold PBST (137 mM NaCl, 2.68 mM KCl, 6.48 mM Na_2HPO_4, 1.47 mM KH_2PO_4, and 0.1% Tween 20, DEPC treated) and left on ice for about 30 minutes. They were then fixed overnight at 4°C on a rocking table with 4% paraformaldehyde in PBST in 5 ml screw cap glass vials. After fixation, embryos were washed twice for 5 min. each with ice cold PBST and then dehydrated stepwise through a graded Methanol/PBST series (25%, 50%, 75%) into 100% methanol. Embryos were stored at -20°C for up to eight months.

2. *Preparation of digoxigenin-labeled RNA probes:*

Approximately 2 µg of the desired plasmid was digested with enzymes cutting at either the T7 or T3 end of the multiple cloning site. After digestion, proteinase K was added at a concentration of 10 mg/ml and incubated for 30 minutes at 37°C. The reaction was then extracted with phenol-chloroform (1:1 v) and the DNA was precipitated by the addition of 2x volume 95% ethanol and resuspended in DEPC-treated water. The *in vitro* transcription reaction was performed from either the T3 or T7 end to obtain sense and antisense RNA probes. The labeling reaction was performed according to the manufacturer's instruction (Boehringer Mannheim). The unincorporated digoxigenin 11-UTP was removed by passing the reaction through a sephadex G-50 (Sigma) column and the RNA was then ethanol precipitated, dissolved in DEPC-treated water and run on a 0.8% agarose gel in order to confirm the integrity and length of the Probe.

I used a Pax1 RNA probe as a positive control (Deutsch et al., 1988). A plasmid carrying a 0.9 kb insert from the Pax1 gene cloned into pBluescript II KS, was cut with *Xba*I and subjected to *in vitro* transcription with the T3 primer, as explained. The Pax1 plasmid was kindly provided by Dr. D. Rancourt, University of Calgary.

3. *Incorporation efficiency*

If necessary, the incorporation efficiency of the digoxigenin-UTP was measured as follows: A serial dilution from the riboprobe was prepared from 10^{-2} to 10^{-5}. 1 µl of each dilution was spotted onto a Hybond nylon membrane filter, and the membrane was baked for 1-2 hr. The filter was preblocked in 1% Blocking reagent (Boehringer Mannheim) in solution 1 (100 mM Tris.HCl, pH 7.5, 150 mM NaCl) for 30 min at RT. The filter was then washed in solution 1 for 15 min, and incubated with alkaline phosphatase-conjugated antidigoxigenein antibody (Boehringer Mannheim) at 1/5000 dilution in solution 1 for 30 min at RT. It was washed with two changes of solution 1 for 30 min, and finally the color detection was performed with NBT/BCIP in 1x alkaline phophatase buffer (100 mM

Tris.Cl, pH 9.5, 50 mM $MgCl_2$, 100 mM NaCl, 0.1% Tween 20). Detection of color in the 10^{-4} dilution indicated successful digoxigenin labeling.

4. *Hybridization of RNA probes to embryos*

Embryos were rehydrated through a graded Methanol/PBST series (75%, 50%, 25%) for 5-10 min. at each step. Then they were washed twice in 100% PBST. At this stage the heart and head of the embryos older than 9.5 days were punctured using fine forceps or needles, in order to prevent the probe from being trapped in these cavities. The embryos were incubated in 4:1 PBST: 30% H_2O_2 for 1 hr, then washed in PBST three times for 5 min each. Proteinase K was added at a final concentration of 5 mg/ml in 0.5 ml of PBST at 25°C. The embryos were treated with proteinase K for one minute at RT. In my hands, the embryos treated longer than 1 min were overdigested.

5. *Removing nonspecific hybridization*

The embryos were washed in solution I (50% formamide, 5x SSC, 0.1% Tween 20) at 70°C three times each for 30 min. The next washes were as follows: 10 min in 1:1 solutionI/solutionII, 15 min wash in several changes of solutionII (0.5 M NaCl, 10 mM Tris.Cl, pH 7.5, 0.1% tween 20) at RT, 1 hr wash at 37°C with two changes of solutionII containing 100 mg/ml RNase A and 100 u/ml RNase T1 (both from Sigma). Embryos were then washed for 1.5 hr at 65°C with three changes of solution III (50% formamide, 2x SSC), and finally washed with 1x TBST (1.4 M NaCl, 27 mM KCl, 25 mM Tris.HCl pH 7.6, 0.1% tween 20) three times for a total of 15 min.

6. *Antibody detection and visualization of the signals*

Embryos were preblocked in 0.5 ml 1x TBST + 10% heat treated goat serum (Sigma)+ 0.5 mg/ml levamisole (Sigma), and incubated at RT for 2.5 hr with shaking. The goat serum was heated for 30 min. at 55°C and stored at -20°C. Levamisole was added in order to block the endogenous alkaline phosphatase activity (Hogan et al., 1996). Blocking solution was replaced with 1x TBST/10% serum containing a 1/5000 dilution of subtracted antibody (alkaline-phosphatase-conjugated sheep anti-digoxigenin Fab fragment (Boehringer Mannheim), and incubated at 4°C overnight. The next day embryos were washed 3x 5 min with 1x TBST + 0.5 mg/ml levamisole and then 6x 1 hr at RT to remove the excess antibody. Finally they were washed for 2x 20 min with freshly prepared phosphatase (AP) buffer (100 mM Tris.Cl, pH 9.5, 50 mM $MgCl_2$, 100 mM NaCl, 0.1% tween 20, 0.5 mg/ml levamisole). The last wash was replaced with 1 ml of AP buffer containing 4.5 ml of NBT (75 mg/ml in 70% DMF, dimethylformamide, BRL) and 3.5 ml of BCIP (50 mg/ml in 100% DMF, BRL).

The embryos were left at RT in the dark for 30 min to 6 hr, depending on the expression level of the genes, until the signal was visible. Finally the embryos were examined and photographed under a fiber optic dissecting microscope (Zeiss).

7. *Preparation of acetone powder and subtraction of antibody*

A newborn mouse or an 18 day fetus was homogenized in PBS, 1 ml per each gram of tissue, and the suspension was transferred to ice for 5 min. 8 ml of cold acetone powder (-20°C) was added for each 2 ml of the cell suspension and mixed vigorously. This was then incubated on ice for 30 min with occasional mixing. The precipitate was collected by centrifugation at 10,000g for 10 min and the supernatant was discarded. The pellet was resuspended in fresh acetone (-20°C), mixed, and placed on ice for 10 min. The suspension was pelleted again and transferred to a clean piece of 3MM Whatman paper. The pellet was spread, dispersed, and let dry at RT. The powder was transferred to an airtight container and stored at -20°C. Any large pieces which did not break into powder were removed.

To subtract the antibody (the Fab fragment from an antidigoxigenin antibody from sheep, conjugated with alkaline phosphatase) the powder was added to a 1:100 dilution of antibody, prepared in TBST+10% sheep or goat serum, at a final concentration of about 1%. It was then incubated for 30 min at 4°C, spun at 10,000g for 10 min, and finally the supernatant was taken and stored at 4°C. An adequate amount (typically to a final concentration of 1:5000) of the subtracted antibody was used for the detection of the digoxigenin-labeled hybridized RNA probe.

Fluorescence in situ hybridization

1. *Preparation of metaphase cells and slides*

Lymphoblast or fibroblast cells from normal and translocation-bearing cell lines (table 2-3) were cultured as explained before. Colcemid was added to a final concentration of 0.2 μg/ml, 1-5 hr before harvest, to arrest the cells at metaphase. The cells were then pelleted, resuspended in 0.075 M KCl, incubated at 37°C for 10 min, and re-pelleted. The

Table 2-3: Translocation cell lines used in FISH analysis

PATIENT CELL LINE	TRANSLOCATION	PHENOTYPE	CELL LINE SOURCE
GM07511A	t(14;22)(q32;q11.2)	Gardner syndrome, multiple polyosis, Osteomas, dental anomalies*	NIGMS
GM03197	t(17;22)(p13;q11) +inv(5)	normal	NIGMS
GM06943	t(21;22)(p12;q11.1)	normal	NIGMS
TOH**	t(21;22)(p12;q11.2)	abnormal facies similar to VCFS patients, mental retardation*	Dr. B.S.Emanuel
AT	t(11;22)(q23;q11)	holoprosencephaly, developmental delay*	Dr. P. Ferreira
BE	t(14;22)(q31;q11)	normal	Dr. M. Nordenskjold

* Only the main abnormalities are listed in this table.
** For some FISH studies a fibroblast culture, GM01700, of this individual obtained from NIGMS was used.
NIGMS: NIGMS Human Genetic Mutant Cell Repository

cells were resuspended in cold fixative (3:1 methanol: glacial acetic acid) by slowly adding the fixative to the cells. The cells were washed twice in fixative, resuspended in cold fixative and stored at -20°C until needed. From 2 to 4 drops of the fixed cells were dropped on a wet, cold slide from 50-75 cm height, and the slides were allowed to air dry and then heated over a 70°C water bath. The slides were typically used within 48 hr.

2. *Preparation of biotin-labeled probes*

Cosmid or phage clones were digested with *Sau*3A and biotin-labeled utilizing a non-isotopic nick translation kit (Oncor or BRL). For most probes approximately 100 ng of digested DNA was used. The probes were prehybridized with 50-100x excess of Cot1 DNA (BRL) to suppress the repetitive sequences before addition to the slides. Cot1 DNA is human placental DNA which is enriched for repetitive sequences such as the *Alu* I and *Kpn* I families (Britten et al. 1986)

3. *Hybridization, washing, and visualization of the signal*

Fluorescence in situ hybridization (FISH) was performed according to the Oncor instruction manual with some modifications. Slides were treated with 70 µl of RNase A (10 mg/ml) and denatured at 70°C in 70% formamide, 2x SSC. Hybridization with biotin-labeled probes was performed in 50% formamide, 2x SSC, 10% dextran sulfate at 37°C overnight. Slides were then washed with shaking at 43°C in 50% formamide/1x SSC, 1x SSC, and 0.1x SSC, respectively, before detection with FITC-labeled avidin. Chromosomes were visualized by counter staining with propidium iodide/antifade (Oncor Inc.). From 25 to 35 metaphase spreads were randomly chosen and examined for each probe. Slides were observed using a Zeiss Axiophot photomicroscope with a standard FITC filter.

Chapter III: Results

The results are presented in three separate sections as follows: Part A: Cloning within the CESCR and isolation of exons in the region, Part B: Chracterization of the putative genes in the CESCR, and Part C: Molecular localization of translocations in proximal 22q11.

Part A: Cloning within the CESCR and isolation of exons in the region

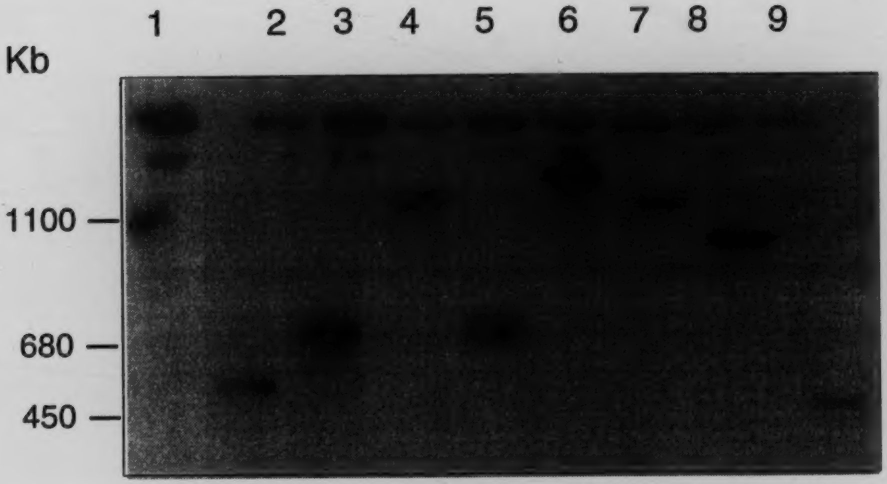
1. A YAC contig within the CESCR

Yeast artificial chromosomes (YACs) in the cat eye syndrome critical region (CESCR) were kindly provided by Dr. C. Bell (Chromosome 22 Genome Center, University of Pennsylvania). An initial contig had been produced using PCR screening of the CEPH YAC library with STSs in the region. My role in this project was to confirm the presence of contiguous loci in overlapping YACs and to establish a YAC contig within the CESCR. The yeast strains containing the YACs were cultured and DNA was extracted and used in Southern hybridization to test for the presence on each YAC of 14 loci located throughout the CESCR. The sizes of the YACs were determined by pulsed field gel electrophoresis (PFGE) by myself and D. Shkolny. An example of a PFGE for sizing YACs has been presented in figure 3-1. YACs are generally prone to deletions, and we noticed a difference in size of some of the YACs examined in different experiments (figure 3-2). The YACs were further used in fluorescence in situ hybridization (FISH) to human metaphase chromosome in order to investigate the integrity of each YAC and possible chimerism (McTaggert K., MS thesis, 1997). The FISH results provided evidence for chimerism in the majority of the YACs, since hybridization was detected on other chromosomes. A minimal tiling path of the YACs which demonstrated hybridization only to chromosome 22 by FISH, or had a large chromosome 22 fragment which could be mapped, is presented in figure 3-2.

2. Isolation and analysis of putative genes

In order to isolate exons from the CESCR, an exon amplification approach was used. A diagram demonstrating an overview of this technique is provided in figure 2-1 in chapter II.

Figure 3-1: Sizing of YACs by PFGE. YAC plugs were electrophoresed through a 1% agarose gel (lane 1-9). The resulting blot was probed with α^{32}P-dCTP- labeled total human genomic DNA. The sizes of the YACs were estimated using yeast chromosomes as markers. Lane 1-9 are YACs: 829D11, 891F12, 925G12, 89C8C1, 805D10, 800A4, 745G7, 878D3, 776H2. Subsequent studies (McDermid et al., 1996) showed that all of these YACs are chimeric. Some of the YACs examined here are shown in figure 3-2.


1
2
3
4
5
6
7
8
9
Kb
1100
680
450

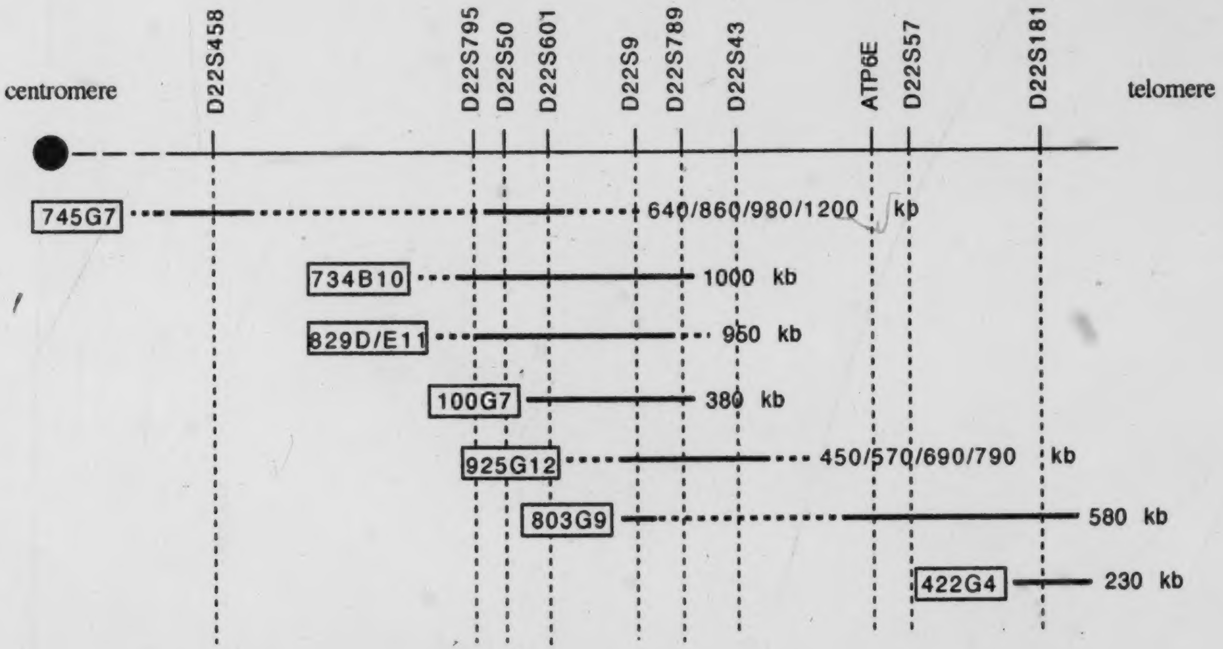

centromere
D22S458
D22S795
D22S50
D22S601
D22S9
D22S789
D22S43
ATP6E
D22S57
D22S181
telomere
745G7
640/860/980/1200 kb
734B10
1000 kb
829D/E11
950 kb
100G7
380 kb
925G12
450/570/690/790 kb
803G9
580 kb
422G4
230 kb

Figure 3-2: A YAC contig map of CESCR. This map is redrawn from McDermid et al. (1996). Loci tested by hybridization against the YACs are shown above the chromosome, with spacing approximately to scale based on a PFGE map. Sizes of the YACs have been determined by PFGE. Some YACs showed various sizes in a single isolate (745G7, 925G12) indicating the instability of these YACs. Dotted lines at the end of YACs indicate possible chimerism, dotted lines within YACs indicate deletions. YAC 829D/E11 was used for isolation of cosmids from a chromosome 22-only cosmid library by Dr. C. Bell. Some of these cosmids were used for the isolation of genes.

Figure 3-3 demonstrates clones in the provisional CESCR which were exon trapped. The region encompasses approximately 1 Mb from D22S795 to D22S57. BACs and PACs were kindly provided by Dr. N. Shimizu (Keio University), Dr. B. Roe (Oklahoma University), and Dr. S. Scherer (Hospital for Sick Children, Toronto) and mapped by Angela Johnson (unpublished results). Some of the cosmids isolated from the chromosome 22-only library in our laboratory or by others have been shown in this figure and are listed in table 3-1. Overall, 2 BACs, 3 PACs and 27 cosmid clones (some overlapping) were used for exon trapping. The clones which resulted in at least one putative exon as well as the number and size of the exons from each clone have been listed in table 3-1. Generally BACs, PACs, and cosmids were digested with *Sst*I and then subcloned randomly into the *Sst*I digested exon trapping plasmids, pSPL3, or pSPL3B. The DNA isolated from these clones were used for electroporation and the subsequent isolation and amplification of the exons. Alternatively, for exon trapping from BACs (b95A8, and b233A), and PACs (p109L3, p120N18, and p238M15) the DNA from these clones was digested with *Sst*I and shotgun cloned into *Sst*I digested pSPL3B and plated on LB+ampicillin plates. Colony lifted filters of these clones were screened with a probe made from the total human genomic DNA and therefore only clones carrying inserts derived from the human genome had DNA prepared for the isolation of exons. This was done to alleviate the problems of the amplification of fragments derived from the cosmid vector or other possible DNA contaminants. On average about 60% of the plated clones hybridized to the labeled total human DNA. After transformation of COS-1 cells and RT-PCR amplification and subcloning of exons, a second source of contamination was splicing of the vector intron due to cryptic splice sites. These clones were eliminated by electrophoresing the isolated exons on a 1.5% gel and hybridizing with a probe made from the intron region of the HIV *tat* gene in pSPL3B. This probe was made by digesting the pSPL3B plasmid with *Hin*dIII and *Sal*I which releases the whole intron leaving only 9 bp of the flanking region. Hybridization with this probe typically resulted in the exclusion of about 20% of the amplified fragments (resulted from vector splicing) when pSPL3B was used as the vector for exon trapping. In my experience the frequency of amplified fragments originated from vector splicing was approximately 20% higher when pSPL3 was used. Approximately 25% of the trapped fragments contained sequences of ALU or L1 repetitive sequences. Rarely, other types of DNA contamination such as fragments from bacterial genomic DNA or the cosmid vector were trapped.

The putative exons were sequenced, and the sequences were run through BLASTN and BLASTP searches. The similarity of the exons to known sequences have been described individually below.

Figure 3-3: A diagram of the clones used in exon trapping and the gene segments identified. The distance of the markers is drawn roughly to scale. Several BACs, PACs, and cosmids (shown with blue boxes) were used in this study. The cosmids c62G1, and c85C7 were isolated by hybridizing the YAC 829D11 to a chromosome 22-only cosmid library (Bell C, unpublished results). The locations of these cosmids have been estimated from the location of the YAC and are not accurate. The putative exons, and genes/pseudogenes isolated from these clones have been shown under each clone. No previously untrapped exons were produced from p120N18 or b95A8. The putative exons shown with asterisks are now currently being sequenced and characterized by P. Brinkman. The orientations of the genes are not known except IDGFL which is shown on the figure.

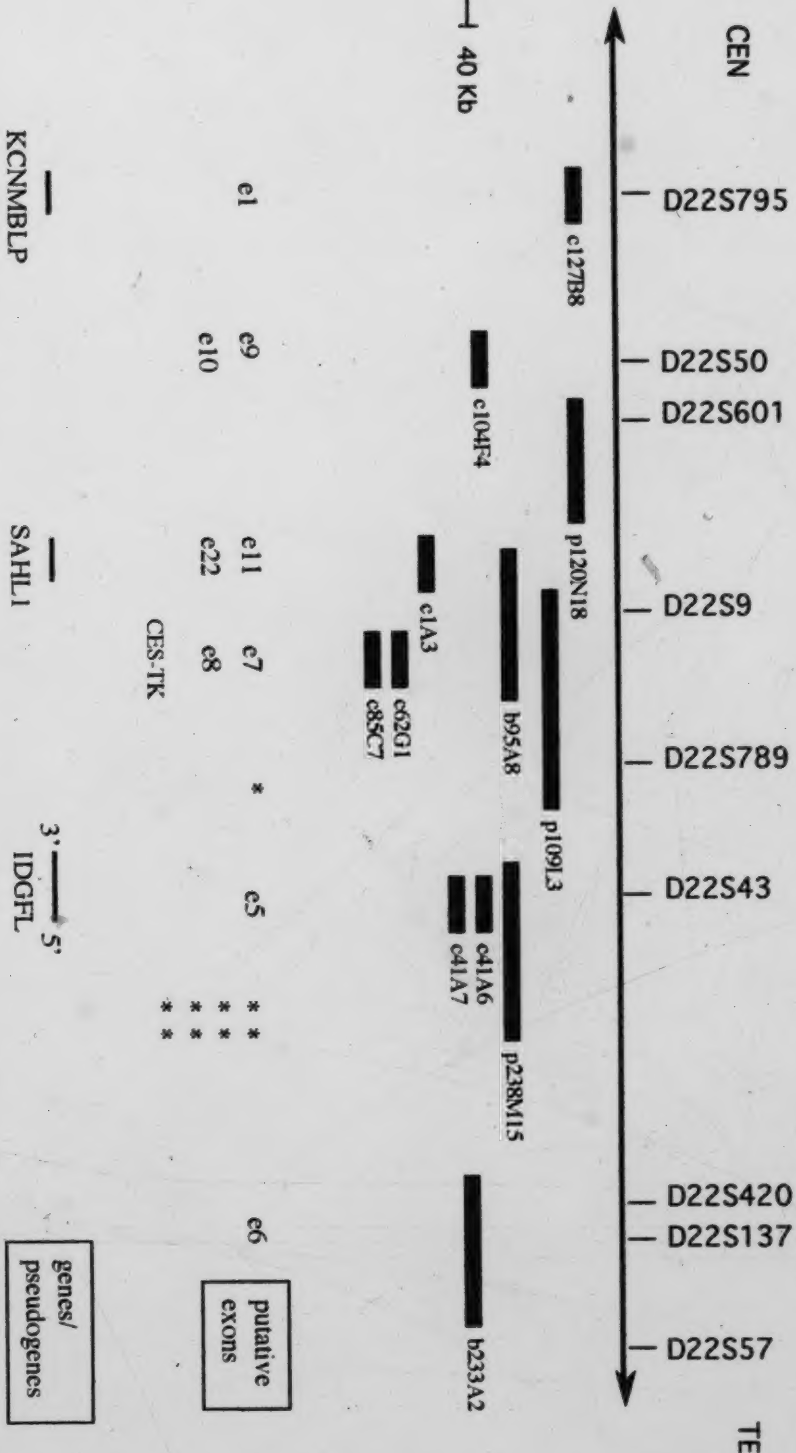

CEN
TEL
D22S795
D22S50
D22S601
D22S9
D22S789
D22S43
D22S420
D22S137
D22S57
40 Kb
c127B8
c104F4
p120N18
b95A8
c1A3
p109L3
c62G1
c85C7
c41A6
c41A7
p238M15
b233A2
e1
e9
e10
e11
e22
e7
e8
*
e5
**
**
**
e6
CES-TK
putative exons
KCNMBLP
SAHL1
3'
5'
IDGFL
genes/ pseudogenes

Table 3-1: The putative exons amplified from CESCR using exon-trapping approach

clone	source	associated locus	putative exon	size (bp)
c127B8	K. McTaggart	D22S601	e1	199
c1A3/c1C7	Dr. N. Shimizu	D22S9	e11	103
			e22	128
c41A6/c41A7	A. Johnson	D22S43	e5	365
c85C7	Dr. C. Bell	D22S9	e7	129
			e8	207
c104F4	K. McTaggart	D22S50	e9	129
			e10	nd
b233A2	Dr. N. Shimizu	D22S420-S137	e6	125
p109L3	Dr. S. Scherer	D22S9-S789	*1 exon	nd
p238M15	Dr. S. Scherer	D22S43	*8 exons	nd

* these exons are being sequenced and characterized by P. Brinkmann
- nd: not determined
-cosmids, BACs, and PACs are indicated with a c, b, and p respectively before the clone address

Part B: Characterization of putative genes in the CESCR

1. Cloning and characterization of pseudogene KCNMBLP, in the vicinity of D22S795, and gene KCNMBL on chromosome 3

Exon isolation and 3' end Cloning

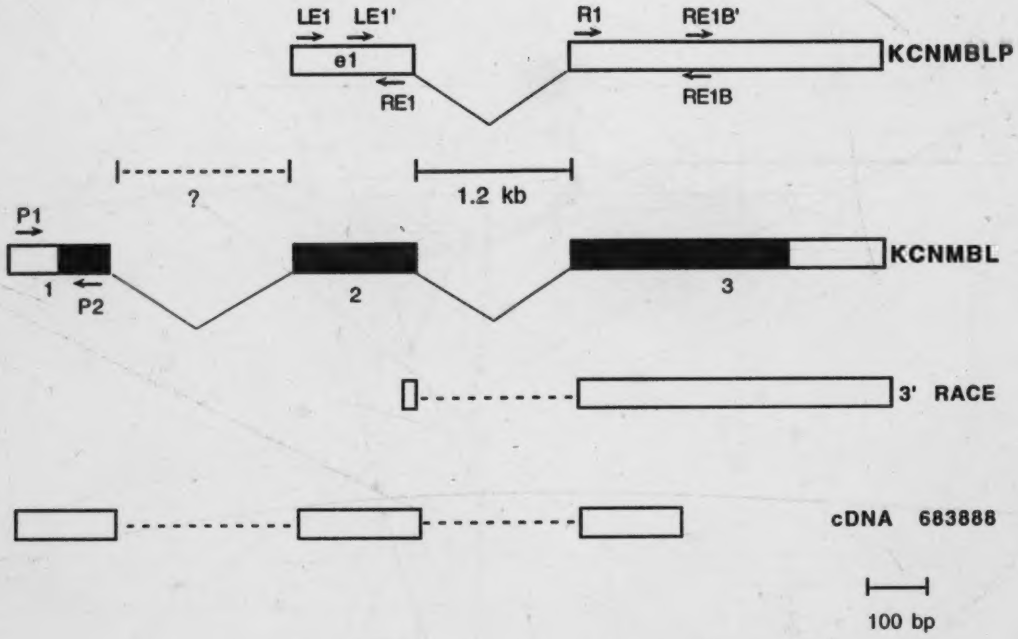
Putative exon "e1" (figure 3-3) was originally isolated from 127B8, a cosmid positive for a probe identifying locus D22S795. This putative exon was used to screen a fetal brain cDNA library (Stratagene). This screening resulted in a 1.8 kb cDNA which was subjected to further characterization and sequencing. The analysis of the partial sequence of this cDNA showed no open reading frame (ORF) as well as the existence of an Alu-like repetitive sequence in the middle of this cDNA (data not shown). The cDNA was assumed to be chimeric, or of an unknown origin, and further sequencing of this cDNA was stopped. The sequence of e1 (which was later found to be exon 2 of the gene) was further extended by 3' RACE utilizing LE1 and LE1' nested primers (figure 3-4). An approximately 0.6 kb cDNA fragment was amplified which hybridized to cosmid 127B8, confirming its mapping to the CESCR (data not shown). However further analysis determined that while e1 was part of a truncated pseudogene on chromosome 22, the 3' RACE product was specific to a gene on chromosome 3.

Isolation of a cDNA by computer sequence search

The sequence of the exons 2 and 3 (figure 3-4) was used to search for similarity to previously identified sequences in databases by performing BLASTN, and BLASTN/dbest searches. An EST cDNA, 683888, was identified that matched the sequence of these exons, and turned out to be from a transcript on chromosome 3. Upon resequencing of this EST cDNA it was found that it contained exon 2 and part of exon 3 as well as a putative first exon. The lack of the last 362 bp of exon 3 (produced by 3' RACE) in this cDNA could be the result of an internal start of the reverse transcription reaction when this cDNA was constructed. A sequence combining the cDNA (683888) and the 3' RACE product was further subjected to computer searches. This putative gene has a 681 bp ORF and codes for a protein with 226 residues. A BEAUTY search demonstrated that this protein has significant similarity (p-value 3.9×10^{-33}) to a regulatory subunit of potassium channels (CO6) in quail (*Coturnix coturnix*) as well as similarity to the beta subunit of the Ca^{+2}-

Figure 3-4: KCNMBL and KCNMBLP genomic structure and sequence. **A:** The diagram represents the genomic structure of KCNMBLP and KCNMBL. The fragment amplified by 3' RACE and cDNA 683888 are also shown in this diagram. A 1.2 kb intron separates the putative exons 2 and 3 of KCNMBLP as determined from the genomic sequence of the region. An intron with a similar size should also account for KCNMBL as discovered by PCR analysis using primers LE1' and RE1B. A fairly large intron probably separates exons 1 and 2 of KCNMBL as the DNA between these exons was unamplifiable using primers P1 and RE1 in my hands. The other primers used in this study for either sequencing or PCR of KCNMBL and KCNMBLP are shown in their approximate position on this diagram. All primers except P1 and P2 amplified both KCNMBL and KCNMBLP. The coding region of KCNMBL is shaded. **B:** The sequence of KCNMBL was derived from combining the sequence of cDNA 683888 and the 3' RACE product. The putative KCNMBL gene is 862 bp long and is composed of three exons. The exon/exon junctions are shown by arrows. The asterisk indicates the stop codon for translation. No canonical polyadenylation signal was identified. The closest sequence (aaactaa) to the consensus poly A adenylation signal sequence is underlined.

A



B

```
atggagacccactagatgtgcacaagaggctgccatccagtactggagaggaccgagccgtg       70
ATG CTG GGG TTT GCC ATG ATG GGC TTC TCA GTC CTA ATG TTC TTC TTG     118
CTC GGA ACA ACC ATT CTA AAG CCT TTT ATG CTC AGC ATT CAG AGA GAA     166
                                              ↑
GAA TCG ACC TGC ACT GCC ATC CAC ACA GAT ATC ATG GAC GAC TGG CTG     214
GAC TGT GCC TTC ACC TGT GGT GTG CAC TGC CAC GGT CAG GGG AAG TAC     262
CCA TGT CTT CAG GTG TTT GTG AAC CTC AGC CAT CCA GGT CAG AAA GCT     310
CTC CTA CAT TAT AAT GAA GAG GCT GTC CAG ATA AAT CCC AAG TGC TTT     358
                                                       ↑
TAC ACA CCT AAG TGC CAC CAA GAT AGA AAT GAT TTG CTC AAC AGT GCT     406
CTG GAC ATA AAA GAA TTC TTC GAT CAC AAA AAT GGA ACT CCC TTT TCA     454
TGC TTC TAC AGT CCA GCC AGC CAA TCT GAA GAT GTC ATT CTT ATA AAA     502
AAG TAT GAC CAA ATG GCT ATC TTC CAC TGT TTA TTT TGG CCT TCA CTG     550
ACT CTG CTA GGT GGT GCC CTG ATT GTT GGC ATG GTG AGA TTA ACA CAA     598
CAC CTG TCC TTA CTG TGT GAA AAA TAT AGC ACT GTA GTC AGA GAT GAG     646
GTA GGT GGA AAA GTA CCT TAT ATA GAA CAG CAT CAG TTC AAA CTG TGC     694
ATT ATG AGG AGG AGC AAA GGA AGG AGG AGA GAA ATC TTA AGA CGG TGG     742
CCA AAT TAA agggctggccttcagatgtctgtgatttctgcaactgaggacc            794
        *
taattatgcctgtctgcaaactaacaatgtaaaacgtaatgattaaagtatcat             847
atttcatgtgggaaaaatttaaaaaaaaaaaaaaaaaa                              882
```

activated potassium channels in human (hslo-beta, or KCNMB) and bovine (P-value of $5.3x\ 10^{-25}$ and $1.3x\ 10^{-24}$). On the basis of this similarity, we have named this gene KCNMBL for KCNMB like. Figure 3-5 demonstrates a protein sequence comparison of KCNMBL with CO6 and hslo-beta.

Southern analysis of KCNMBL

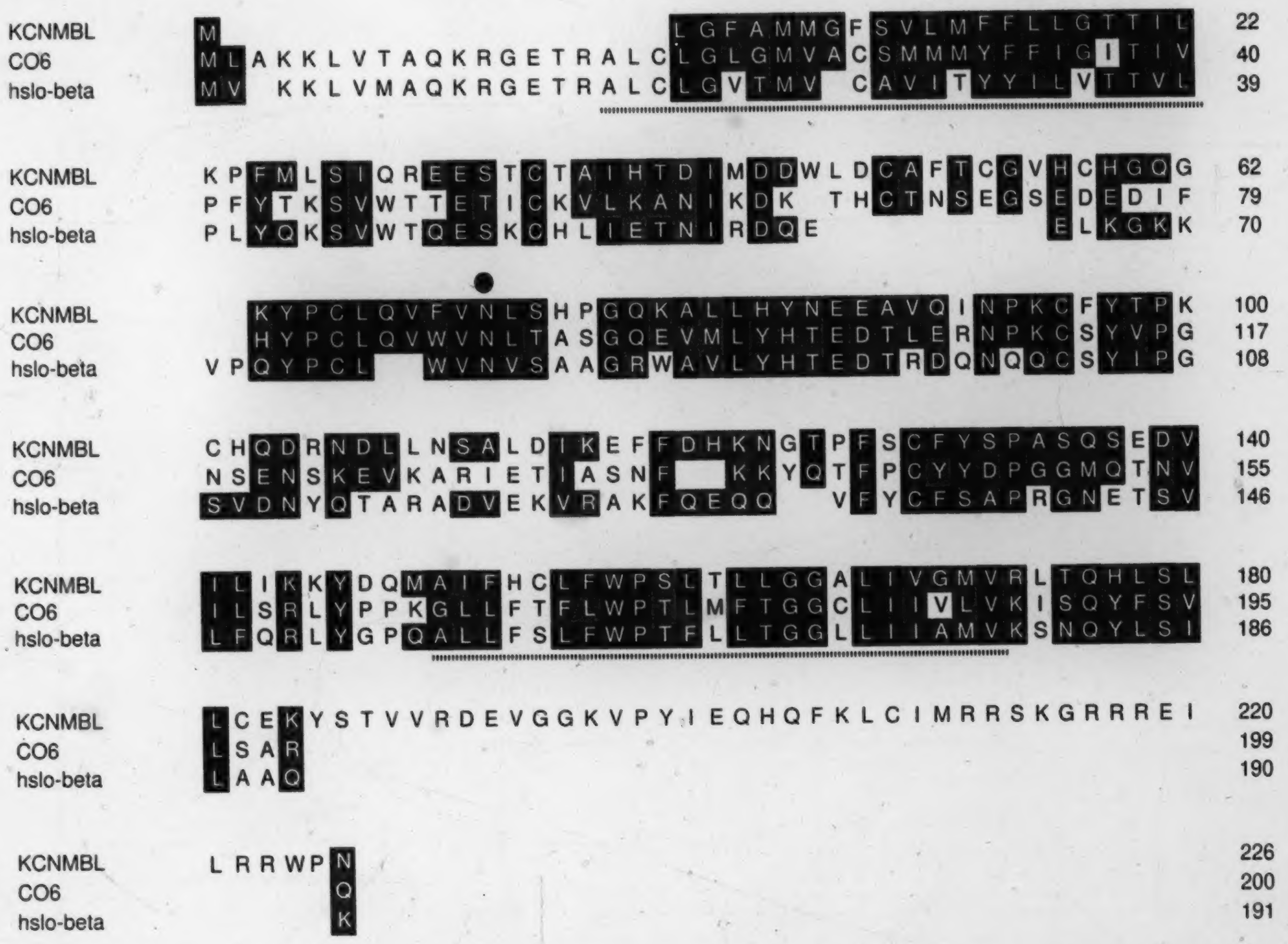
Southern hybridization was undertaken to determine whether the chromosome 22-derived exon was unique or present in multiple copies. The original hybridization of e1 to *Hin*dIII digested genomic DNA from a monochromosome hybrid panel revealed a second copy of this exon on chromosome 3 (not shown). Therefore further DNA studies were done only with the hybrids specific for chromosome 22 and 3. Hybridization of the *Eco*RI and *Pst*I digested DNAs from these hybrids to the labeled 3' RACE detected different-sized bands for both (figure 3-6 A). However when a probe made from exon 1 of KCNMBL by PCR amplification (using primers P1 and P2 indicated in figure 3-4) was hybridized to DNA from these hybrids only a band on chromosome 3 was detected (figure 3-6 B).

Hybridization of the 3' RACE product to DNA from other organisms revealed one copy of this gene in the monkey as well as the chicken genome. A single weak band in rat genomic DNA was detected, while no apparent hybridization was obtained in the mouse genomic DNA (figure 3-6 C), although the latter is probably due to poor quality of the mouse genomic DNA used.

Sequence comparison of KCNMBL with the sequence of the putative exons on b7G2 in the CESCR

The genomic sequence of a chromosome 22-derived BAC (b7G2) in the vicinity of D22S795, provided by Dr. N. Shimizu (Keio University), was examined to determine the exon/intron junction of the gene segment in 22q11. The sequence of exons 2 and 3 on this BAC did not match perfectly with the sequence of the cDNA, suggesting that KCNMBL is not on this BAC. The sequence comparison revealed 97.4% identity between the sequence of exon 2 and 3 of KCNMBL (which localizes to chromosome 3) and the gene segment on chromosome 22 (figure 3-7). A 1219 bp intron separates exons 2 and 3 on b7G2. This was also confirmed with PCR using primers LE1' and RE1B (figure 3-4). PCR amplification of a chromosome 3-only hybrid cell line (GM10253A) using these primers produced a product of a similar size, suggesting a highly conserved genomic structure of KCNMBL on chromosome 3 and the sequence on chromosome 22. Attempts to amplify the exon 1 of

Figure 3-5: The comparison of the deduced amino acid sequence of KCNMBL with CO6 and hslo-beta proteins. The sequences were aligned using the program ALIGN (Genestream resource centre, URL: http://genome.eerie.Fr/bin/align-guess.cgi). The identical amino acids are highlighted with black, and the similar residues with gray. KCNMBL demonstrates 23.8% and 22.9% identity with CO6 and hslo-beta respectively. The identity is higher (54.2% and 56.5%) over the two putative transmembrane domains which have been underlined. The conserved N-glycosylation region of the protein is indicated by a dot.



```
KCNMBL     M                  LGFAMMGFSVLMFFLLGTTIL   22
CO6        MLAKKLVTAQKRGETRALCLGLGMVACSMMMYFFIGITIV   40
hslo-beta  MV KKLVMAQKRGETRALCLGVTMV CAVITYYILVTTVL   39

KCNMBL     KPFMLSIQREESTCTAIHTDIMDDWLDCAFTCGVHCHGQG   62
CO6        PFYTKSVWTTETICKVLKANIKDK THCTNSEGSEDEDIF   79
hslo-beta  PLYQKSVWTQESKCHLIETNIRDQE         ELKGKK   70

                      ●
KCNMBL       KYPCLQVFVNLSHPGQKALLHYNEEAVQINPKCFYTPK  100
CO6          HYPCLQVWVNLTASGQEVMLYHTEDTLERNPKCSYVPG  117
hslo-beta  VPQYPCL  WVNVSAAGRWAVLYHTEDTRDQNQQCSYIPG  108

KCNMBL     CHQDRNDLLNSALDIKEFFDHKNGTPFSCFYSPASQSEDV  140
CO6        NSENSKEVKARIETIASNF  KKYQTFPCYYDPGGMQTNV  155
hslo-beta  SVDNYQTARADVEKVRAKFQEQQ  VFYCFSAPRGNETSV  146

KCNMBL     ILIKKYDQMAIFHCLFWPSLTLLGGALIVGMVRLTQHLSL  180
CO6        ILSRLYPPKGLLFTFLWPTLMFTGGCLIIVLVKISQYFSV  195
hslo-beta  LFQRLYGPQALLFSLFWPTFLLTGGLLIIAMVKSNQYLSI  186

KCNMBL     LCEKYSTVVRDEVGGKVPYIEQHQFKLCIMRRSKGRRREI  220
CO6        LSAR                                      199
hslo-beta  LAAQ                                      190

KCNMBL     LRRWPN                                    226
CO6             Q                                    200
hslo-beta       K                                    191
```

Figure 3-6: Southern analysis of KCNMBL. **A:** Genomic DNA from a normal individual (GM03657 lanes 1, and 4), chromosome 22-only hybrid cell line (GM10888, lanes 2, and 5), and chromosome 3-only hybrid cell line (GM10235A, lanes 3, and 6) was digested with *Eco*RI (lanes 1, 2, and 3) or *Pst*I (lanes 4, 5, and 6) and electrophoresed. A Southern blot of this gel was hybridized to the 3' RACE product, which resulted in hybridization to both chromosome 22, and chromosome 3. The faint bands in lanes 2, and 3 are the result of cross-hybridization to hamster DNA. **B:** Genomic DNA from GM03657 (lane 1), GM10888 (lane 2), GM10235A (lane 3), and genomic DNA from a hamster cell line (lane 4) were digested with *Pst*I and hybridized to a probe prepared from exon 1 of KCNMBL. Hybridization of one fragment in total genomic DNA was seen also in the hybrid for chromosome 3. **C:** Genomic DNA of GM10888 (1), GM03657 (2), mouse (3), rat (4), rhesus monkey (5), and chicken (6) was digested with *Eco*RI and hybridized with a probe made from the 3' RACE product. Only one apparent band is seen in chicken and rhesus monkey. The band detected in the rat genomic DNA is very faint. In addition no discrete band is detected in the mouse genomic DNA. The hybridization was performed at 60°C as explained in chapter II.

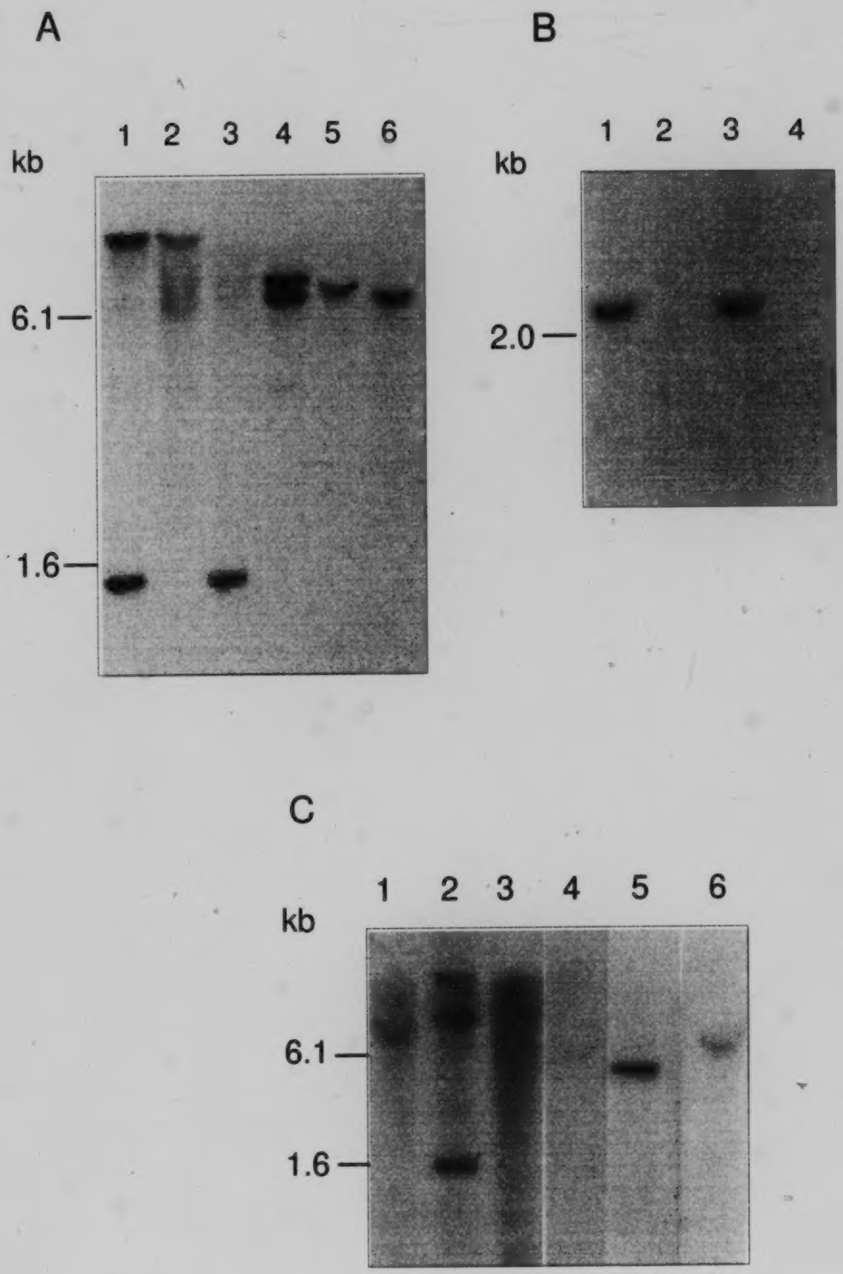

A
B
kb
1 2 3 4 5 6
6.1
1.6
kb
1 2 3 4
2.0
C
kb
1 2 3 4 5 6
6.1
1.6

Figure 3-7: Sequence comparison between KCNMBL on chromosome 3 and KCNMBLP on chromosome 22. The sequence of KCNMBL is indicated by arrows. The two sequences are 97.4% identical. The different nucleotides are shown by vertical lines.

```
  CATTCGGAGAGAAGAATCAACCTGCACTGCCATCCACACAGATATCATGGACGACTGGCT
  :::::|::::::::::::|:::::::::::::::::::::::::::::::::::::::::
→ CATTCAGAGAGAAGAATCGACCTGCACTGCCATCCACACAGATATCATGGACGACTGGCT

  GGACTGTGCCTTCACCTGTGGTGTGCACTGCCGTGGTCACGGGAAGTACTCGTGTCTTCG
  ::::::::::::::::::::::::::::::::||:::::|:::::::::|:|:::::::|
→ GGACTGTGCCTTCACCTGTGGTGTGCACTGCCACGGTCAGGGGAAGTACCCATGTCTTCA

  GGTGTTTGTGAACCTCACCCATTCAGCTCAGAAAGCTCTCCTACATTATAATGAAGAGGC
  :::::::::::::::::|::::|:::|:::::::::::::::::::::::::::::::::
→ GGTGTTTGTGAACCTCAGCCATCCAGGTCAGAAAGCTCTCCTACATTATAATGAAGAGGC

  TGTCCAGATAAATCCCAAGTGCTTTTACACACCTAAGTGCCACCAAGATAGAAATGATTT
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
→ TGTCCAGATAAATCCCAAGTGCTTTTACACACCTAAGTGCCACCAAGATAGAAATGATTT

  GCTCAACAGTGCTCTGGACATAAAAGAATTCTTCGATCACAAAAATGGAACCCCCTTTTC
  :::::::::::::::::::::::::::::::::::::::::::::::::::|::::::::
→ GCTCAACAGTGCTCTGGACATAAAAGAATTCTTCGATCACAAAAATGGAACTCCCTTTTC

  ATGCTTCTACAGTCCAGCCAGCCAATCTGAAGATGTCATTCTTATAAAAAAGTATGACCA
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
→ ATGCTTCTACAGTCCAGCCAGCCAATCTGAAGATGTCATTCTTATAAAAAAGTATGACCA

  AATGGCTATCTTCCACTGTTTATTTTGGCCTTCACTGACTCTGCTAGGTGGTGCCCTGAT
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
→ AATGGCTATCTTCCACTGTTTATTTTGGCCTTCACTGACTCTGCTAGGTGGTGCCCTGAT

  TGTTGGCATGGTGAGATTAACACAACACCTGTCCTTACTGTGTGAAAAATATAGCACTGT
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
→ TGTTGGCATGGTGAGATTAACACAACACCTGTCCTTACTGTGTGAAAAATATAGCACTGT

  AGTCAGAGATGAGGTAGGTGGAAAAGTACCTTATATAGAACAGCATCAGTTCAAACTGTG
  ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
→ AGTCAGAGATGAGGTAGGTGGAAAAGTACCTTATATAGAACAGCATCAGTTCAAACTGTG

  CAGTAGGAGGAGGAGCAAAGGAAGGAGCAGAGAAATCTTAAGACGGTGGCCAAATTAAAG
  ::|::|:::::::::::::::::::::|::::::::::::::::::::::::::::::::
→ CATTATGAGGAGGAGCAAAGGAAGGAGGAGAGAAATCTTAAGACGGTGGCCAAATTAAAG

  GGCTGGCCTTCAGATGTCTGTGATTTCTGCAACTGAGGACCTAATTATGCCTGTCTGCAA
  :|::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
→ GTCTGGCCTTCAGATGTCTGTGATTTCTGCAACTGAGGACCTAATTATGCCTGTCTGCAA

  ACTAACAATGTAAAACGTAATGATTAAAGTATCATATTTTCATGTGGGAAAAAATTT
  :::::|:::::::::|:::::|:::::::::::::::::::::::::::::::::::
→ ACTAATAATGTAAAAGGTAATAATTAAAGTATCATATTTTCATGTGGGAAAAAATTT
```

KCNMBL from b7G2 or a number of overlapping clones (BACs: b566C8, b779H12, and YACs: 745G7, and 734B10) failed. The gene segment on this BAC which is composed of two exons, highly similar to exons 2 and 3 of KCNMBL on chromosome 3, has been designated as KCNMBLP for KCNMBL pseudogene.

Study of the expression of KCNMBL

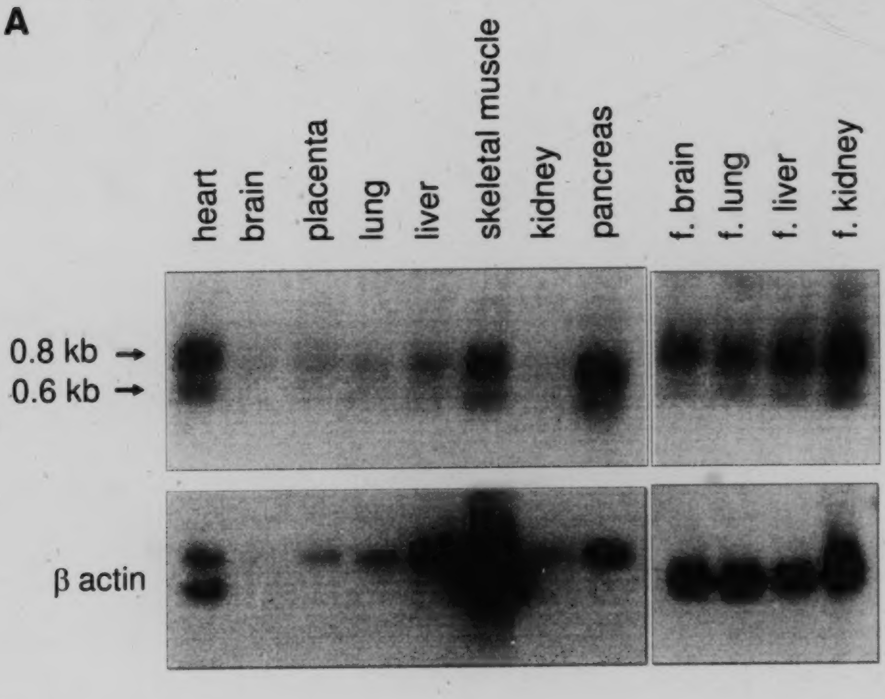
In order to investigate the expression of this gene, Northern blots containing RNAs from adult and fetal tissues were hybridized with a probe made from exon 3. This hybridization revealed an approximately ~0.8 kb and a fainter ~0.6 kb transcripts. Both transcripts are widely expressed in the tissues examined (figure 3-8 A).

RT-PCR was done to confirm the expression of exon 1 as part of this transcript. This was performed on total RNA extracted from a 9 week-old fetus using the primers P1 and RE1 (figure 3-4) and generated a product of the correct size (figure 3-8 B).

2. Partial cloning and characterization of a gene, SAHL1, adjacent to D22S9

Two exons of this putative gene, exons e11 and e22, were isolated using the exon trapping method from a pool of cosmids c1A3, c1H9, c1C7, c88G3 (overlapping clones, only c1A3 has been shown in figure 3-3) kindly provided by Dr. N. Shimizu. These exons hybridized back to the cosmids c1A3, c1H9, and c1C7. Hybridization of e22 to normal human DNA (GM03657) and DNA extracted from a chromosome 22-only hybrid cell line (GM010888) confirmed its mapping to only chromosome 22 (figure 3-9 A). A Zoo blot of the *Eco*RI digested genomic DNA from other organisms was also probed with e22 and revealed unique bands in monkey, chicken, rat, mouse and hamster (figure 3-9 A and B). Database searches of sequences similar to e11 and e22 resulted in an EST cDNA 23249 with perfect identity with these exons. This clone contained a 397 bp insert and was resequenced to confirm the sequence in the EST database. The sequence of this cDNA is presented in the appendix. The cDNA has 56% DNA identity to the 3' end of a previously identified gene, SAH (Samani et al., 1994), over the region of e11 and e22. The rat counterpart of this gene has been suggested to be overexpressed in the kidney of a hypertensive rat strain (Iwai and Inagami, 1991). On the basis of this homology this gene was called SAHL1 for SAH like 1 and this cDNA is refered to as SAHL1-1.

Figure 3-8: Expression of KCNMBL. **A:** Northern blots containing RNA from adult and fetal tissues (fetal RNA is indicated by f.) hybridized with a probe made from exon 3 of KCNMBL. The lower panel demonstrates probing with β–actin which serves as loading control. The lower band seen in the heart and skeletal muscle is due to the cross-hybridization of the β-actin probe to α-actin mRNA. **B:** RT-PCR was performed using primers in exon 1 and 2 (P1, and RE1 respectively). "+" indicates the RT-PCR reaction and "–" indicates the control reaction, in which all steps were done except no reverse transcriptase was added. A product is only amplified in the "+" lane indicating the expression of exon 1 as part of the transcript.


A
heart
brain
placenta
lung
liver
skeletal muscle
kidney
pancreas
f. brain
f. lung
f. liver
f. kidney
0.8 kb →
0.6 kb →
β actin

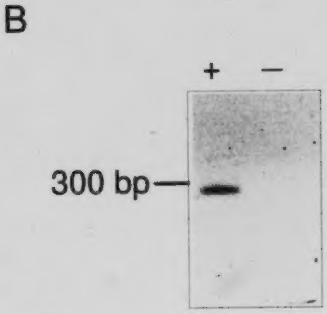
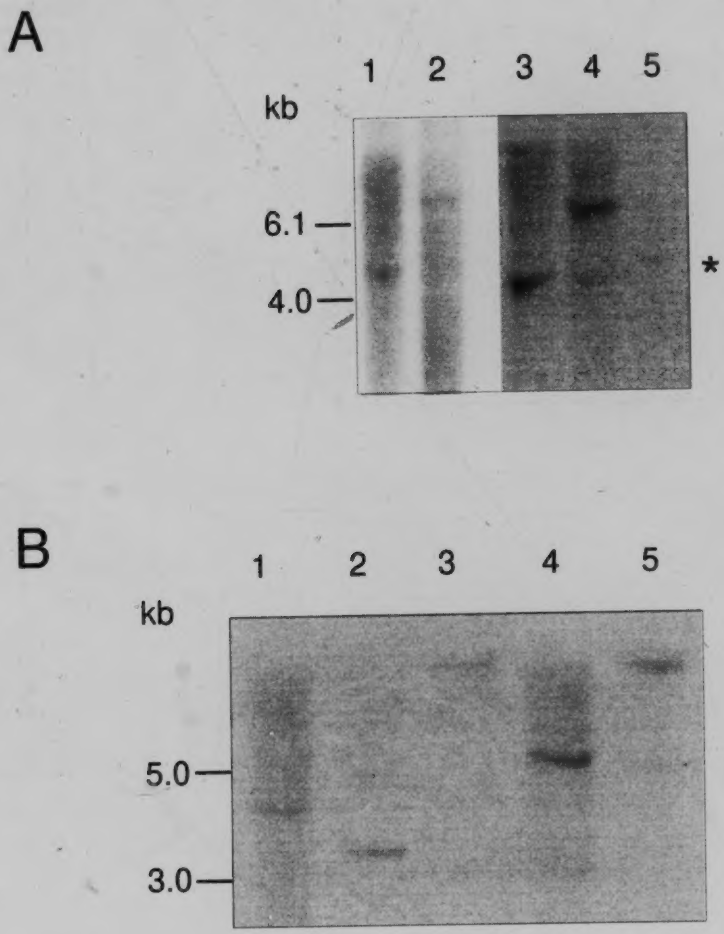

B
+
–
300 bp

Figure 3-9: Southern and Zoo blot analysis of e22. **A:** Approximately 5 µg of genomic DNA from a normal individual, GM03657, (1, 2), and a chromosome 22-only hybrid cell line, GM10888, (3, 4) were digested with *Sst*I (1, 3), and *Pst*I (2, 4). The Southern blot was hybridized with labeled e22. The result suggest that the putative exon e22 is localized to chromosome 22. The faint band seen in lane 5 (mouse genomic DNA) is labeled with an asterisk. **B:** a Zoo blot was prepared from genomic DNA of rat (1), monkey (2), chicken (3), hamster (4), and human (5). The blot was hybridized to e22 at 55°C, and exposed for a week. Single bands are seen in all the lanes.


A
1
2
3
4
5
kb
6.1
4.0
*
B
1
2
3
4
5
kb
5.0
3.0

Northern hybridization of SAHL1-1 to human and mouse RNA

Human adult and fetal Northern blots were hybridized with a 0.3 kb probe made from the SAHL1-1 cDNA by digesting with *Bam*HI. This fragment detected two transcripts with approximate sizes of 2.1 and 2.8 kb expressed only in liver and kidney (figure 3-10). A similar pattern of expression was observed on a Northern blot prepared from mouse tissues (figure 3-10). The human fragment was hybridized to the mouse Northern blot at the same stringency as the human Northern blots, suggesting a high degree of conservation between the human and the mouse genes.

Identification of two cDNAs from a Caco-2 cDNA library

Two cDNAs with the size of approximately 2.1 and 2.6 kb were isolated from a Caco-2 cDNA library. These cDNAs (shown in figure 3-11 as SAHL1-2 and SAHL1-3) were mapped by restriction enzymes and partially sequenced. The sequence analysis revealed the sequence of e11 and e22 as well as some of the flanking exons. Further sequencing of SAHL12 indicated that this cDNA contained only part of exon a (figure 3-11). As well, there was a run of 35 (T)s in the middle of this cDNAs. Therefore, it was assumed that this cDNA might be chimeric. This was confirmed by hybridizing a 0.3 kb *Eco*RI/*Sst*I fragment of this cDNA (from region c in figure 3-11) to a Northern blot which resulted in a different sized transcript. Northern hybridization with a 0.6 kb *Pst*I/*Sst*I fragment from region f of SAHL1-3 (figure 3-11) resulted in an approximately 3 kb transcript with widespread expression (not shown), indicating that this cDNA was also chimeric. Consequently further sequencing of SAHL1-2 and SAHL1-3 was stopped.

Analysis of the cosmid c1A3 sequence reveals partial genomic structure of SAHL1-1, 2, and 3

The genomic sequence of the cosmid c1A3 was kindly provided by Dr. N. Shimizu. The sequence was analysed by Grail-2 and GAIA to predict exons, ESTs, or other landmarks in this sequence. The result revealed a high frequency of repetitive sequences and only a few predicted exons (figure 3-11). Grail-2 analysis of this sequence predicted the exons e11 and e22. The sequence of c1A3 also matches three ESTs: 32483, 41a1, and FB20C7. The EST 32483 is in the same direction as SAHL1-1, and therefore, could represent the 3' end of the chimeric cDNA SAHL1-3 (figure 3-11). Comparing the sequences of SAHL1-1, as well as the region of SAHL1-2 and SAHL1-3 present on c1A3,

Figure 3-10: Expression of SAHL1 in human and mouse. **A:** Human adult and fetal Northern blots were hybridized with *Bam*HI-digested 0.3 kb fragment from the SAHL1-1 cDNA. Two transcripts (2.1, and 2.8 kb) were detected in adult and fetal kidney and liver. **B:** A Northern blot prepared from RNAs from different adult mouse tissues as well as a whole 18 day fetus, probed with the SAHL1-1 fragment. At least two transcripts are detected in kidney and liver. Hybridization was performed at the same condition for the human Northern blots.

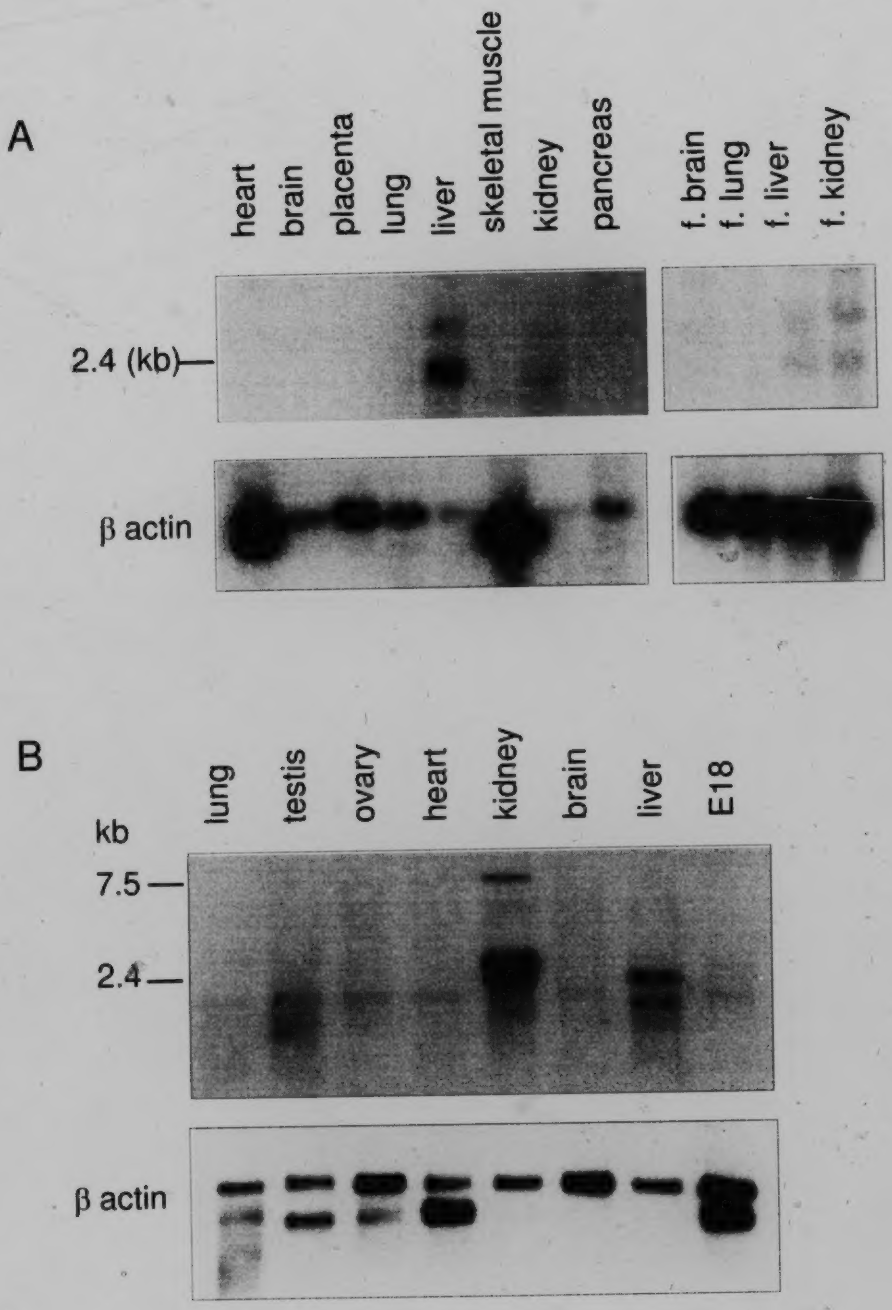
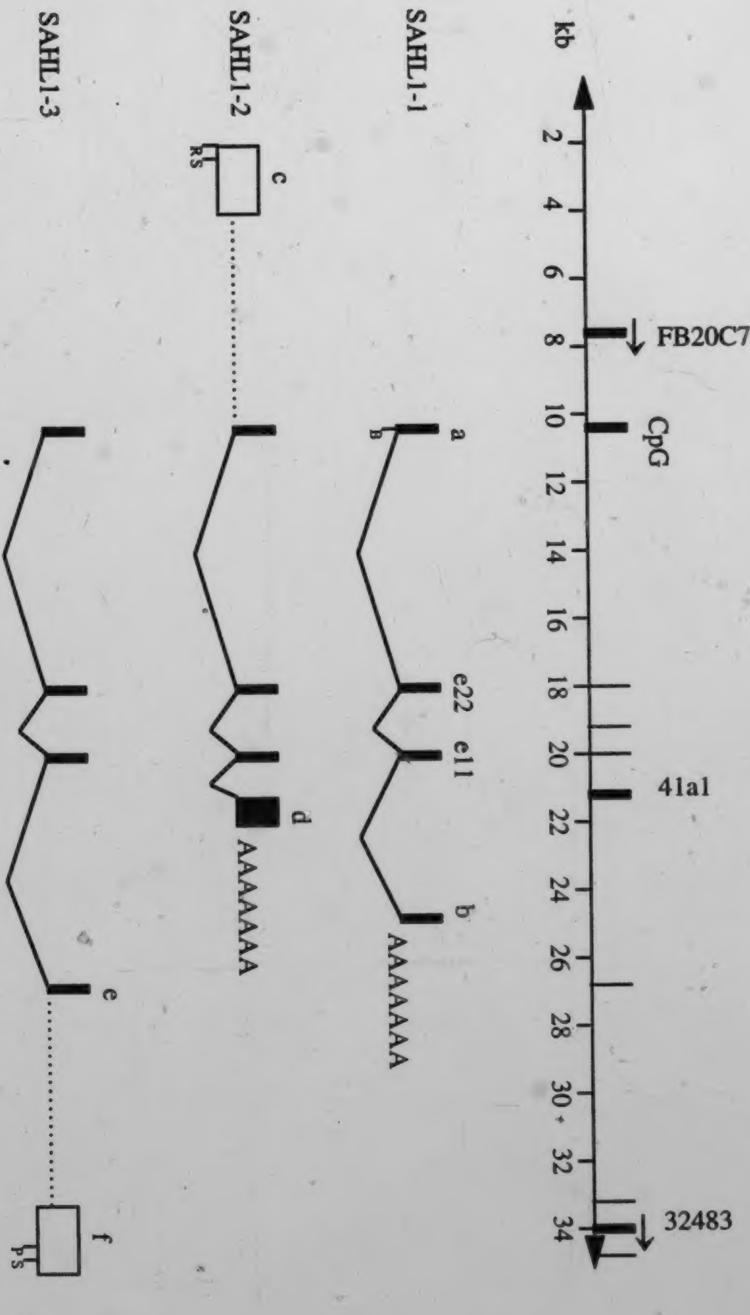

A
heart
brain
placenta
lung
liver
skeletal muscle
kidney
pancreas
f. brain
f. lung
f. liver
f. kidney
2.4 (kb)
β actin
B
lung
testis
ovary
heart
kidney
brain
liver
E18
kb
7.5
2.4
β actin

Figure 3-11: Alternative splicing and partial genomic structure of SAHL1. The sequence of a cosmid, c1A3, in this region has been searched for the presence of coding sequences. The exons, ESTs, and a CpG island identified by Grail-2 and GAIA analysis have been indicated on the sequence. Putative Grail-2 exons are shown with lines, while the ESTs and CpG island are indicated with filled boxes. The transcription orientation of the ESTs are shown with an arrow. The sequence of EST FB20C7 on c1A3 is the 3' end of the clone, while the sequence of EST 32483 on this cosmid is the 5' end of this clone. A comparison between partial sequence of SAHL1-1, 2, and 3, and the genomic sequence has revealed the exon/intron boundaries. The run of (A)s indicates the 3' end of the cDNAs. Fragments c and f of SAHL1-2 and SAHL1-3 were not detected on this sequence and proved to be derived from unrelated cDNAs. The position of the restriction enzymes: R (*Eco*RI), S (*Sst*I), P (*Pst*I), and B (*Bam*HI), which were used for production of the probes has been shown.


kb
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
34
FB20C7
CpG
41a1
32483
SAHL1-1
SAHL1-2
SAHL1-3
a
b
c
d
e
f
e22
e11
AAAAAAA
AAAAAAA
RS
PS

with that of the cosmid reveals the partial genomic structure of these genes. This also resulted in identification of two more exons "d" in SAHL1-2 and "e" in SAHL1-3 (figure 3-11). The exon "d" overlaps with EST 41a1. This EST has been isolated from a retinal cDNA library using random primers, however, it was not available for further characterization. FB20C7 is located 5' to the SAHL1-1, 2, and 3 sequence on c1A3, therefore, there is a possibility that it represents the same gene, however, RT-PCR using primers within this cDNA and SAHL1-1 (Fb-2 and 11-2, table 2-2) failed to detect any product (result not shown). In addition a probe made from the PCR amplification of FB20C7 using Fb-1 and Fb-2 primers (table 2-2) did not detect the same transcript as SAHL1-1, but a very faint band >4 kb in size (not shown). The sequence of the putative exons of SAHL1-1, 2, and 3 has been presented in the appendix. In order to amplify the missing 5' portion of these genes I attempted 5' RACE using total RNA extracted from a 9 week-old fetus, however, this was not successful.

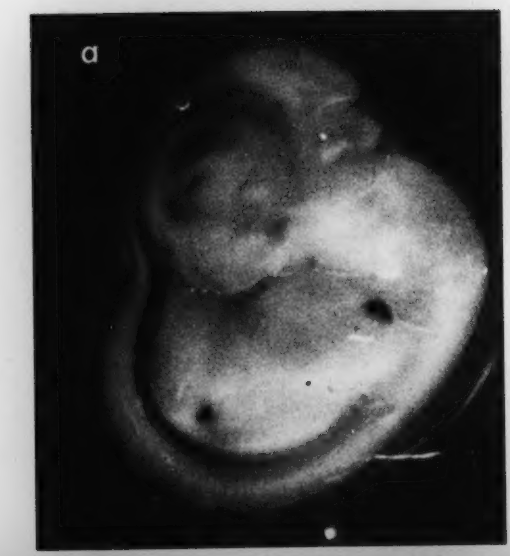
Mouse embryonic expression of SAHL1-1

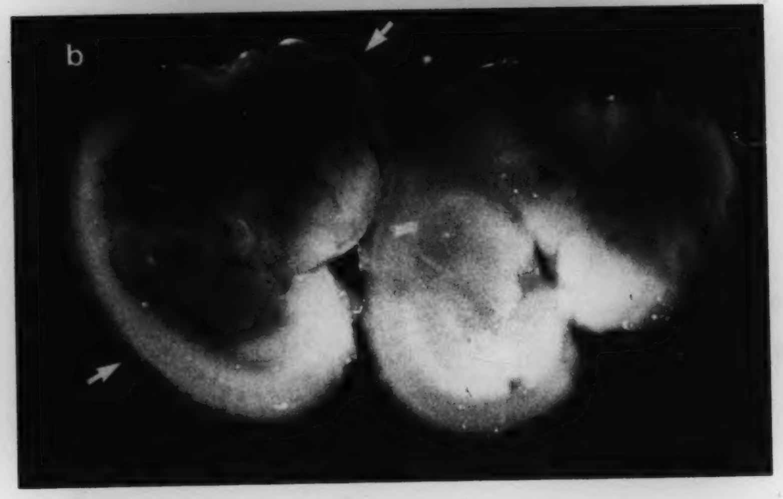
The embryonic expression of SAHL1-1 was investigated by whole mount mouse *in situ* hybridization. The embryos were obtained from mating of the CD1 parents. Digoxigenin-labeled sense (negative control) and antisense (test probe) riboprobes were prepared from SAHL1-1 and used to probe the 11.5 dpc embryos. An RNA probe made from the mouse Pax-1 gene was used as a positive control in all the experiments. The preliminary result suggests expression in the region of developing kidney in 11.5 dpc embryos (figure 3-12).

Isolation of a mouse BAC for SAHL1-1 to search for a region of the mouse genome homologous to the CESCR

I used a probe made from SAHL1-1 to screen a BAC library made from the genomic DNA of mouse (kindly provided by Dr. B. Birren). The screening resulted in a positive clone. The BAC clone is now being mapped to mouse metaphase chromosomes in Dr. J. Korenberg's laboratory.

Figure 3-12: Whole mount *in situ* hybridization of SAHL1. A riboprobe prepared from pPAX1HIS plasmid was used as positive control (a), which demonstrates hybridization to the somites. Sense and antisense digoxigenin-labeled riboprobes made from the SAHL1-1 cDNA were hybridized to 11.5 dpc embryos (b). The arrows indicate the suspected hybridization of the antisense probe to the region of developing kidney in the embryo.


a


b

3. Cloning and characterization of a gene, IDGFL, in the vicinity of D22S43

A 356 bp exon, e5, was first isolated by exon trapping from a pool of cosmids 41A6 and 41A7 in the region of D22S43. PCR of the genomic DNA using primers, B-1, and B-2 (table 2-2) amplified a 1.5 kb product suggesting that e5 is in fact composed of two exons (result not shown). This putative exon was also localized by Southern hybridization to BACs and PACs and showed hybridization to PAC p238M15 (figure 3-3). Southern hybridization using e5 to human genomic as well as a chromosome 22-only hybrid DNA (GM10888) suggested a single copy of this gene in the CESCR (figure 3-13).

Identification of IDGFL cDNA and sequence analysis

Searching of the EST database produced a matching cDNA, 54445, which was purchased from TIGR (The Institute for Genomic Research). The size of the cDNA (3.7 kb) matched approximately with the size of the transcript detected on the Northern blot (see below), and therefore, this cDNA was further characterized by sequencing. The sequence of this cDNA is presented in figure 3-14. This cDNA has an approximately 1.5 kb ORF encoding a protein with significant similarity (p-value of 10^{-102}) to a putative growth factor from *Sarcophaga peregrina* (IDGF, Homma et al., 1996), as well as to atrial gland granule-specific antigen (AGSA) of *Aplysia californica* (Sossin et al., 1989) (figure 3-15). Consequently I have designated this gene "IDGFL" for insect-derived growth factor like.

Although I attempted to use the whole cDNA sequence as a probe for Southern blotting, the probe constantly demonstrated non-specific hybridization even when prehybridized with total human DNA to block repetitive sequences. This probably resulted because of the abundance of the repetitive sequences within the 3' untranslated region of this gene (figure 3-14). Probing the PAC p238M15 with e5 as well as the cDNA 54445 revealed only one band (1.5 kb) in the *Sst*I digest of this PAC. However, when the cDNA was hybridized to the adjacent more centromeric PAC, p213P2, it detected several bands (result not shown). This suggested that the orientation of this gene is telomeric to centromeric (figure 3-3) with e5 localizing to the most 5' end of the cDNA.

Northern analysis of IDGFL

Exon e5 was used to probe human adult and fetal Northern blots, and detected an approximately 3.5 kb transcript (figure 3-16). This gene is expressed in all the fetal tissues

Figure 3-13: Southern hybridization of e5. Total genomic DNA from GM03657 (lanes 1, and 2), and GM10888 (lanes 3, and 4) were digested with *Sst*I (lanes 1, and 3) and *Pst*I (lanes 2, and 4) and electrophoresed. The resulting blot was hybridized to labeled e5 at high stringency, indicating that there is only one copy of this sequence in the genome.

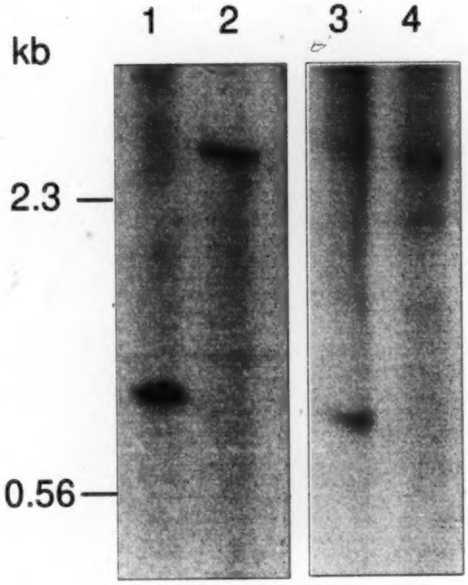
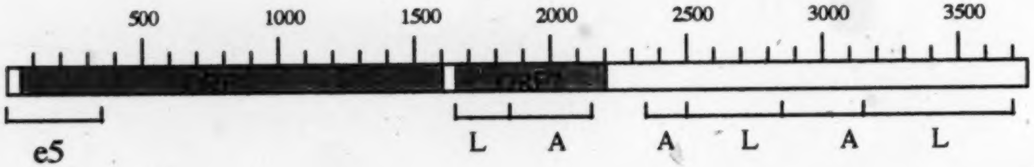
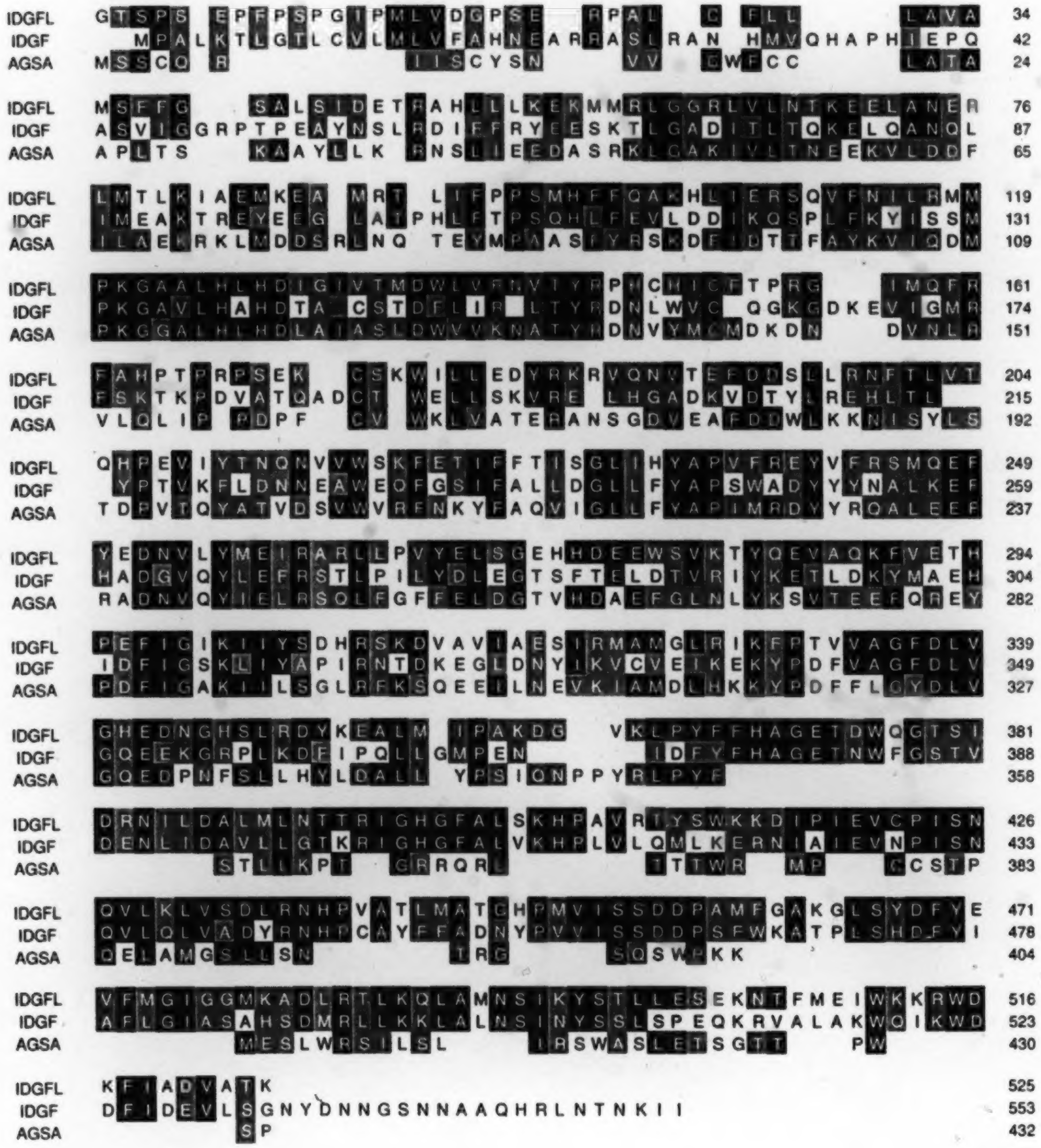

1
2
3
4
kb
2.3
0.56

Figure 3-14: Sequence of the IDGFL cDNA. A diagram of the cDNA and the location of the ORFs is shown above. The first ORF (ORF1) extends from base 43 to 1578 and ORF2 encompasses bases 1666 to 2190. The position of e5 is also shown on the diagram. The regions (bases 1649-1842, 1856-2158, 2380-2496, 2514-2860, 2863-3163, and 3165-3690) indicated by A and L demonstrate a high degree of identity at DNA and/or protein sequence with different subfamilies of Alu repetitive sequence (A) and L2 repeat (L). In the sequence of the IDGFL cDNA, the putative coding sequence (ORF1) is indicated in uppercase letters while the 5' and 3' untranslated regions are in lower case. The consensus polyadenylation signal is underlined.



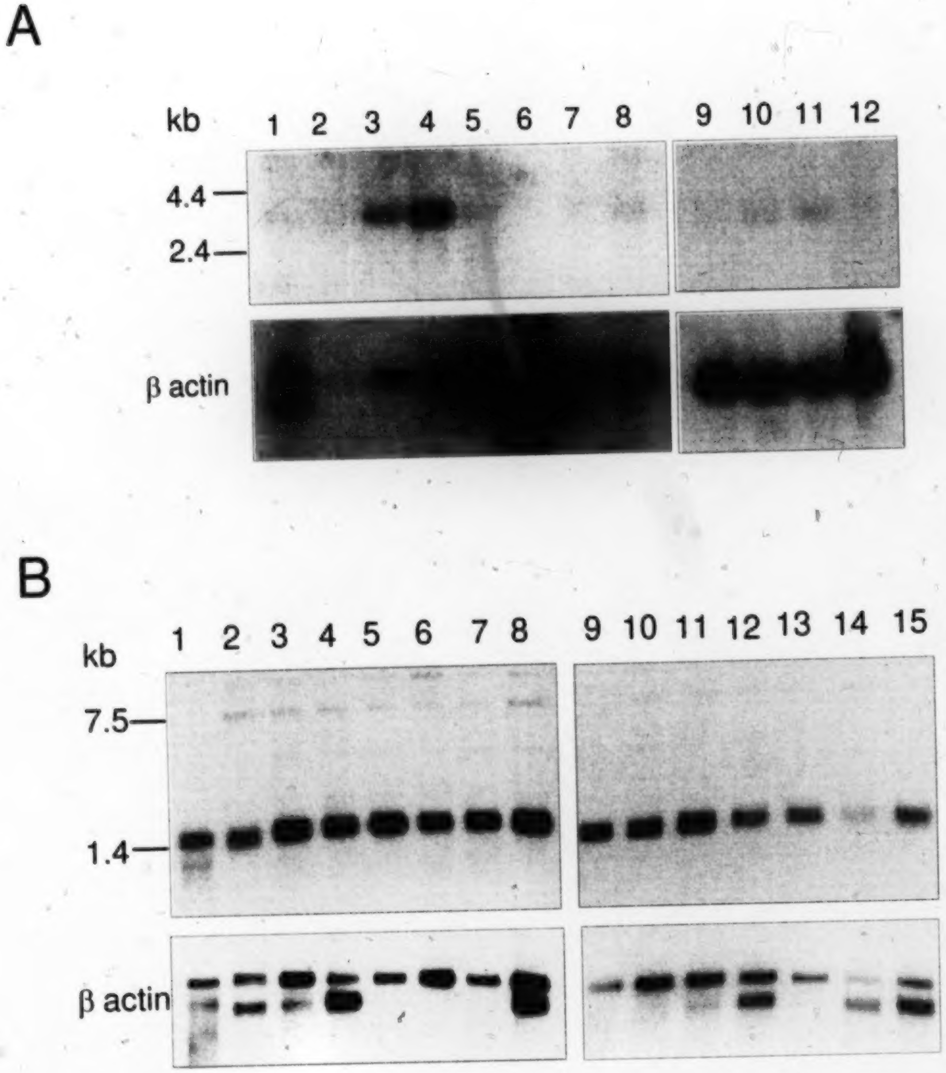
```
ggcacgagtccatctgagccctttcctagtccaggcatcccg ATG TTG GTG GAT GGC CCA TCT GAG CGG CCA GCC CTG TGC TTC TTG CTG TTG GCT GTG GCA 117
ATG TCT TTC TTC GGC TCA GCT CTA TCC ATA GAT GAA ACA CGG GCG CAT CTG TTG TTG AAA GAA AAG ATG ATG CGG CTC GGG GGG CGG CTG GTG 213
CTG AAC ACC AAG GAG GAG CTG GCC AAT GAG AGG CTC ATG ACG CTC AAA ATC GCT GAG ATG AAG GAG GCC ATG AGG ACC CTG ATA TTC CCA CCC 309
AGC ATG CAC TTT TTC CAG GCC AAG CAT CTC ATT GAG AGA AGT CAA GTG TTT AAT ATT CTA AGG ATG ATG CCA AAA GGG GCT GCC TTG CAC CTC 405
CAT GAC ATT GGC ATC GTG ACT ATG GAC TGG CTG GTG AGG AAT GTC ACC TAC AGG CCT CAC TGC CAC ATC TGT TTC ACC CCA AGG GGG ATC ATG 501
CAG TTC AGA TTT GCT CAC CCA ACT CCC CGT CCA TCA GAA AAA TGT TCC AAG TGG ATT CTG CTG GAG GAT TAT CGG AAG CGG GTG CAG AAC GTC 597
ACT GAG TTT GAT GAC AGC TTG CTG AGG AAT TTC ACT CTG GTG ACC CAG CAC CCG GAG GTG ATT TAC ACA AAC CAA AAT GTT GTC TGG TCG AAA 683
TTT GAA ACC ATC TTC TTC ACC ATC TCT GGT CTC ATC CAT TAC GCA CCA GTG TTC AGA GAA TAT GTC TTC CGG AGC ATG CAG GAG TTT TAC GAG 779
GAC AAC GTG CTT TAC ATG GAG ATC AGA GCC AGG CTG CTG CCG GTG TAT GAG CTC AGT GGA GAG CAC CAT GAC GAA GAG TGG TCA GTG AAG ACT 875
TAC CAG GAA GTA GCT CAG AAG TTT GTG GAA ACT CAC CCT GAG TTT ATT GGA ATC AAA ATC ATT TAT TCG GAT CAC AGA TCC AAA GAT GTG GCT 9701
GTC ATC GCA GAA TCC ATC CGA ATG GCC ATG GGG CTC CGA ATC AAG TTC CCC ACG GTG GTG GCA GGG TTT GAC CTG GTG GGG CAT GAG GAC AAT 1017
GGC CAC TCC TTG CGT GAC TAC AAG GAA GCT TTG ATG ATC CCC GCC AAG GAT GGC GTT AAG CTG CCT TAC TTT TTC CAC GCC GGA GAA ACA GAC 1113
TGG CAG GGT ACT TCC ATA GAC AGG AAC ATT TTG GAT GCT CTG ATG CTG AAC ACT ACC AGA ATC GGC CAT GGA TTT GCT TTG AGC AAA CAC CCC 1209
GCA GTC AGG ACT TAC TCC TGG AAA AAG GAC ATC CCC ATA GAA GTC TGT CCC ATC TCT AAC CAG GTG CTG AAA CTG GTG TCT GAC TTG AGG AAC 1305
CAC CCT GTA GCC ACT CTG ATG GCC ACT GGG CAC CCC ATG GTG ATC AGC TCT GAT GAC CCA GCT ATG TTT GGT GCC AAA GGC TTG TCC TAT GAT 1401
TTC TAT GAG GTC TTC ATG GGC ATT GGG GGG ATG AAG GCT GAC CTG AGG ACC CTC AAA CAG CTG GCC ATG AAC TCT ATC AAG TAC AGT ACC CTG 1497
TTG GAG AGT GAG AAA AAT ACT TTC ATG GAA ATC TGG AAG AAG AGA TGG GAT AAG TTC ATA GCA GAT GTG GCT ACA AAG TGA ggagaagctagccag 1593
ccctctacaagctgtcttcttgcacacgctgtcacttcctctcactcgttcttgaatcagctccatgtgcccatgaaatcaatggcctctgtatggagcgaccctgtgagaagcacttggctg 1716
gctgagcaaattcatcctctggaaatattctctctcagccacagtgacattgaccctcttggttttctcctctctctggccatttcttccagtttccctatttcagagtcttctcctctctct 1839
gatctctgtgctgtttcctcaggactcagtcctgggctctcttctattctggtctctttatttttttatttttgtatttttttcgagatggagttttgctcttgttgcccaggctggagtacaa 1962
tggtgcgatctcagctcagtgcaacctccgccacccgggttcaggcaatttctcttgcatcagcctcccgagtagttggaattataggcatgtgccaccacacccagctgatttttgcattttt 2085
agtagagacaggttttcaccatgttggcgagctggtatccaactcttgacctcaggtgatccactcgccccttggctcccaaagtgctggaattacaggcgttagccaccatgcctggcctat 2208
tctggtctctttaactctctcctctttatttctcttctctctctgtacactttttcctgggtggtctcatccattcctttgcttttttcataccatttatttgttaatgattcccacatttattt 2331
atgcacttggagagctcacaggaatctcagaaactgatgaggtacaattctgaaccctcagtctcttccctttaaacctttcttttttctctactttaatttttctaaagagtgtcttgctatg 2454
ttgcccaggctggtctccaactcaagtgatcctcctgccgcagtctcccgaagtgctgggattactgacatgagccaccacactcagccctttaaacctttccctggcctttcccatagctgg 2577
tgaaggacacctccatccattccacgcagttgctcaaagcagaaattttcagtgcaagtcttgatgctgcgccgtcccccactccctacatcagaacgcatccctcatctggactccagcggt 2700
ggcttcttgatgctgcgcggtcccccactccctacatcagaatgcatcccgcatccagactccagcggtggtgctctacctgcacgctgttgccaagtccaagctaccatactcctgcctgag 2823
ctatgacaacagcctcctcactgatctcccctttcttccctttgcctcctccagctcatttttcacagtgtagaatgacattttgtttgtttgttttgttttgtttgagatggagtctcgttc 2947
tgttgcccagggtggagtgcagcggtgcgatctcggctcactgcaacttccactcccgggttcaagcggattctcgtgcctcagcctcctgagtagctgggattacaggcatgcaccaccatg 3069
cccggataatttttgtatttttagtagagatggggtttcactatgttggccaggctggtctcgaacacctgacttcgtggtccacccgccttcggcctcccaaagcactgggattacaggcgt 3192
gagccacccggcctggcctagaatgacttttaaaagatcaaattaaatcaggtcactcctttgcttacaacgcagtgcgtttagaggtacaccccatgtctccacagggcatacagcatccga 3315
tttaatctggatccattccggcgcttcctctcccagtcacccagagggccccaaccccggcggccctttcttcctcaaatgtcctcggctctataccgtgcctgggtcttttctctttctctc 3438
tgcctggaacattccttctttccccttttgtcttgcccactcctgtttacccttcaagtttcaagttcatgtcactgtctcagagaggttttcctgtgctcgccctgtttctctcaggaagcc 3591
ttgctcttttccatcatgcctctaatcacagcttataatcggatatttatttctgtgtctacagtcttgccctgccagactgtaagccccatgtgggcaggcgctcatgattgtttctgattg 3684
tttcacgcatgctgctaacccagagcctgggcccaaagctagttagtactcaataaacaatgcattgaaaaaaaaaaaaaaaaaanta                                       3771
```

Figure 3-15: Protein sequence comparison of IDGFL with IDGF and AGSA proteins. The sequences were aligned using the program ALIGN. The identical amino acids are highlighted in black, and the similar residues with gray. IDGFL demonstrates 34.8% identity with IDGF and 26.7% identity with AGSA.



```
IDGFL GTSPS EPFPSPGIPMLVDGPSE  RPAL  C FLL     LAVA 34
IDGF    MPALKTLGTLCVLMLVFAHNEARRASLRAN HMVQHAPHIEPQ 42
AGSA  MSSCQ R         IISCYSN    VV  GWFCC     LATA 24

IDGFL MSFFG   SALSIDETRAHLLLKEKMMRLGGRLVLNTKEELANER 76
IDGF  ASVIGGRPTPEAYNSLRDIFFRYEESKTLGADITLTQKELQANQL 87
AGSA  APLTS   KAAYLLK RNSLIEEDASRKLGAKIVLTNEEKVLDDF 65

IDGFL LMTLKIAEMKEA MRT LIFPPSMHFFQAKHLIERSQVFNILRMM 119
IDGF  IMEAKTREYEEG LATPHLFTPSQHLFEVLDDIKQSPLFKYISSM 131
AGSA  ILAEKRKLMDDSRLNQ TEYMPAASFYRSKDFIDTTFAYKVIQDM 109

IDGFL PKGAALHLHDIGIVTMDWLVFNVIYRPHCHICFTPRG   IMQFR 161
IDGF  PKGAVLHAHDTA CSTDFLIR[illegible]LTYRDNLWVC QGKGDKEVIGMR 174
AGSA  PKGGALHLHDLAIASLDWVVKNATYRDNVYMCMDKDN   DVNLR 151

IDGFL FAHPTPRPSEK  CSKWILLEDYRKRVQNVTEFDDSLLRNFTLVT 204
IDGF  FSKTKPDVATQADCT WELLSKVRE LHGADKVDTYLREHLTL   215
AGSA  VLQLIP RDPF  CV WKLVATERANSGDVEAFDDWLKKNISYLS 192

IDGFL QHPEVIYTNQNVVWSKFETIFFTISGLIHYAPVFREYVFRSMQEF 249
IDGF   YPTVKFLDNNEAWEQFGSIFALLDGLLFYAPSWADYYYNALKEF 259
AGSA  TDPVTQYATVDSVWVRFNKYFAQVIGLLFYAPIMRDYYRQALEEF 237

IDGFL YEDNVLYMEIRARLLPVYELSGEHHDEEWSVKTYQEVAQKFVETH 294
IDGF  HADGVQYLEFRSTLPILYDLEGTSFTELDTVRIYKETLDKYMAEH 304
AGSA  RADNVQYIELRSQLFGFFELDGTVHDAEFGLNLYKSVTEEFQREY 282

IDGFL PEFIGIKIIYSDHRSKDVAVIAESIRMAMGLRIKFPTVVAGFDLV 339
IDGF  IDFIGSKLIYAPIRNTDKEGLDNYIKVCVEIKEKYPDFVAGFDLV 349
AGSA  PDFIGAKIILSGLRFKSQEEILNEVKIAMDLHKKYPDFFLGYDLV 327

IDGFL GHEDNGHSLRDYKEALM IPAKDG  VKLPYFFHAGETDWQGTSI 381
IDGF  GQEEKGRPLKDFIPQLLGMPEN      IDFYFHAGETNWFGSTV 388
AGSA  GQEDPNFSLLHYLDALL YPSIQNPPYRLPYF             358

IDGFL DRNILDALMLNTTRIGHGFALSKHPAVRTYSWKKDIPIEVCPISN 426
IDGF  DENLIDAVLLGTKRIGHGFALVKHPLVLQMLKERNIAIEVNPISN 433
AGSA        STLLKPT  GRRQRL       TTTWR  MP   CCSTP 383

IDGFL QVLKLVSDLRNHPVATLMATCHPMVISSDDPAMFGAKGLSYDFYE 471
IDGF  QVLQLVADYRNHPCAYFFADNYPVVISSDDPSFWKATPLSHDFYI 478
AGSA  QELAMGSLLSN       TRG     SQSWPKK            404

IDGFL VFMGIGGMKADLRTLKQLAMNSIKYSTLLESEKNTFMEIWKKRWD 516
IDGF  AFLGIASAHSDMRLLKKLALNSINYSSLSPEQKRVALAKWQIKWD 523
AGSA         MESLWRSILSL    IRSWASLETSGTT   PW      430

IDGFL KFIADVATK                                     525
IDGF  DFIDEVLSGNYDNNGSNNAAQHRLNTNKII                553
AGSA         SP                                     432
```

Figure 3-16: Expression analysis of IDGFL and the putative mouse homologue, mC1. **A:** Human adult and fetal Northern blots were hybridized with e5. Lanes 1-8 are loaded with RNA from adult heart, brain, placenta, lung, liver, skeletal muscle, kidney, pancreas; and lanes 9-12 are fetal brain, lung, liver, and kidney. The adult brain lane is underloaded since only a faint transcript is detected when probed with β-actin. **B:** Mouse adult and fetal Northern blots were hybridized with mC1. The lanes 1-8 are RNA from: adult lung, testis, ovary, heart, kidney, brain, liver, E18 (18 day fetus); and the lanes 9-15 are RNA from: E7, E12 head, E12 body, E18 heart, E18 liver, E18 head, E18 except head, liver, and heart. Hybridization with β-actin was performed as an RNA-loading control. Note that the size and expression pattern differs from that of e5 in humans, suggesting that this clone is not the true homologue of IDGFL.


A
kb
1
2
3
4
5
6
7
8
9
10
11
12
4.4
2.4
β actin
B
kb
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
7.5
1.4
β actin

tested, as well as placenta. The adult expression is more specific to the lung; however, weak hybridization to RNA from heart, brain, liver, and pancreas is also detected.

Attempts to isolate a mouse homologue for IDGFL

Exon e5 was also used to screen a mouse fetal cDNA library. This resulted in 6 positive clones of different sizes which were isolated and subcloned into pBluescriptII SK$^+$. Southern hybridization of DNA from these clones with e5 revealed the strongest hybridization to clone mC1. Further, these cDNAs were hybridized with a fragment of the IDGFL cDNA which did not contain e5. This probe also strongly hybridized to mC1 suggesting that the homology extends beyond e5.

Northern blots of mouse RNAs from adult and fetal tissues were hybridized with mC1. This probe detected an approximately 1.4 kb transcript in all the tissues tested (figure 3-16).

The chromosomal location of mC1 was investigated in collaboration with Dr. R. Reeves (Johns Hopkins, Baltimore). The preliminary result has indicated that this clone does not map to the mouse chromosome 16 (which we expected to contain the region with homology to the CESCR (personal communication)). This together with the Northern hybridization result suggests that this cDNA might not be the real mouse counterpart of IDGFL, if such a homology even exists. In addition, the sequence of IDGFL did not detect any mouse EST in the dbEST database.

4. Characterization of e7 and e8

Isolation of the exons and cDNAs

Exons e7 and e8 were isolated from cosmid c87C7 in the area of D22S9 (figure 3-3). The search for homologous sequences for e8 revealed marginal similarity of this sequence to collagen type II (p-value 10^{-5}). Exon e7 was used to screen a fetal brain as well as a Caco-2 cDNA library. This resulted in the isolation of a 5 kb cDNA, c7/8 which was isolated from the fetal brain cDNA library three times in two attempts. This cDNA was hybridized to a Southern blot containing DNA from BACs and PACs in the region and demonstrated strong hybridization to these clones. In addition it showed strong hybridization to both e7 and e8 (not shown). Consequently, it was sequenced in collaboration with Dr. Shimizu's laboratory. Despite the strong hybridization to e7 and e8, the sequence of c7/8 demonstrated 74.2% and 76.5% DNA identity to e7 and e8 respectively (figure 3-17). The largest ORF detected in this 5 kb cDNA encodes 115 amino acid residues. Therefore this cDNA is apparently from a homologue somewhere else in the genome. Southern hybridization of e7 to total genomic DNA revealed strong hybridization to chromosome 22, however, a faint band, not on chromosome 22, is also detected (not shown). The hybridization of the same blot with a 0.9 kb *Sst*I fragment of c7/8, encompassing a region with no DNA similarity to e7 (figure 3-17), strongly detected the faint band seen with e7 hybridization. This confirmed that c7/8 putative cDNA was from another region in the genome. I did not attempt to determine the genomic location of c7/8 in this study.

Isolation of the mouse genomic and cDNA clones for e7

Putative exon e7 was used to isolate mouse genomic clones. A mouse genomic library was screened by A. Johnson with this putative exon which resulted in two phage clones. One of the clones showed a stronger hybridization to the exon. The phage clone (m7B) was sent to Dr. J. Squire (Hospital for Sick Children, Toronto) to be mapped to mouse chromosomes by FISH. The clone showed hybridization to chromosome 16 on the border of bands B and C. In approximately 30% of metaphase spreads hybridization to another chromosome (either 9 or 10) was detected (figure 3-18).

A mouse 8.5 dpc cDNA library was also screened by this probe and resulted in two clones, me7A (0.7 kb), and me7B (~4 kb) which were partially sequenced. The mouse clones demonstrate 96.2% identity with each other. The best local identity between these cDNAs and e7 is 54.7% over 64 nucleotides. Comparison of the DNA sequence of these cDNAs with c7/8 reveals a slightly higher similarity (62.3% DNA identity over 77 nucleotides). Clone me7A also showed strong hybridization to the previously identified genomic clone (m7B). The sequence has been presented in the Appendix.

Search for candidate genes in the region

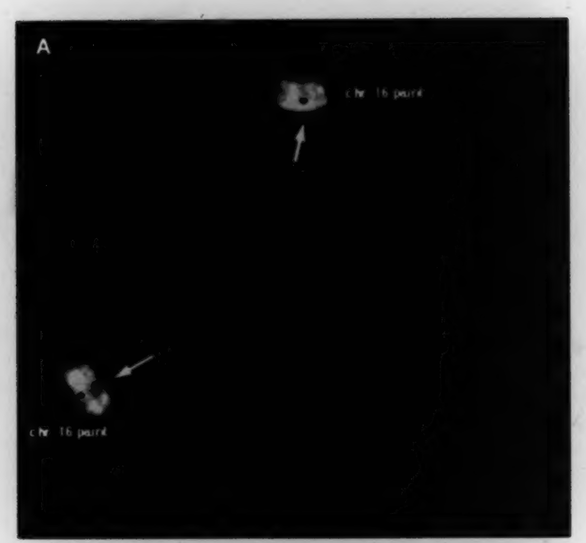
1. *CES-TK*

Several YACs (745G1, 734B10, 800A4, 829D11, 100G7) and a number of cosmid clones in the CESCR were subjected to PCR with degenerate primers for the conserved domain of tyrosine kinase proteins (table 2-2, D'Esposito et al. 1994). PCR using these primers amplified a 266 bp DNA fragment (called CES-TK) from the cosmid 62G1. This cosmid was provided by Dr. C. Bell using hybridization screening of the

Figure 3-17: The sequence of cDNA c7/8. The total size of the cDNA is 5216 bp. The region with similarity to e7 and e8 have been shown by bold characters. A segment of e8 (e8I) shows 72.2% identity with c7/8, while another segment of e8 (e8II) shows 80.9% identity with this cDNA. The percentage of identity with e7 is 74%. The 0.9 kb *Sst*I fragment used in Southern analysis of this cDNA has been underlined.

5'...acttagatgacacaaatttcttgaaagacacaaactaccaaaattcattcaagaaaaaaaaagctgcatccattgtggggaggaagaaggtgggcacgggatctt
cctgagctgtctgaaggggagaagagcatggcctctgcctgccactgaccagcctcaggagcctggcttctctttctgaatgtgggctttaaagctcagactccaaact
catcatccctcctgcaggatgtagcagcacctgccagcccttcttacttcccggcttctgggaggacccaggcccttccttggccctggtggatgggtccaggccttca
tctctgcctccacccatctctgtcaggtcgctcatccagggaaggaatttgctaggtgcaaacaaacatcagccacaggactttacgtgctgtgctcactgccaccctgt
gaaggggaaccatgctgaggcacagctctgcacccttctgtctagactgcttagctataaggatgcctcctggggaggctttcccaggcctggaagtcccttctgaag
gctgatgtatcgacggcagggctctggacccactcgagagcccaggaccaggaggcagtatcctcattcctggatctgttgctgattcactttgtaaccttaggacaat
cagatctctgtgcactcatttgcccatctctacaacacgggggatgggctgactgatgaggttctctgactgagggtactgtgctctgagagcaggaataggagtcttta
gacctagatgaaagctctatgactaaagatgcatcataaaagctgggtactggagcacgagggaggctgagtgctgctgttcagtgttgaaggcaccctggcctccat
gatctctgcctccgggaccctgagagctcttactgccccccaactctactccaggcaccttgcggtgggggcttggggaGCTCGGCTGGAAGTCC
AGGGCCAGATAATCAACGCTGCCGGTGCTCTTCTTAGGGGCAGGACTGTTCGTGCCACTGGGA
e8I ACGGGAGATGCAGACACTGGGTTTTGCTGtcacgaggaggaaaaaactgtgagtgactgcagaaaacagactaacccaccacccca
gaaaaacacctctgggagtaagaacactgtttccctgagccctactctgacacatgcaccatgctaaggacttttcatgttttcattttaatcctcatgacaatcctgtaagg
taggtatgattactgttgatactgtttaaagcaaagatgaaactgtggttcactttactttcagagaagtgactttgcctgtggctgcacagctgttgagtgatgggaatcag
gtctgttttctgacttccaaacctgtgctaacccctgtaccacactgcctcctttctagcacatcaggccaagtctcatcaacagtggctctcctgagacatccagcctttgc
ttgctggcgcttaagaaaggagcacatttcctttatgcactggaaagttttcttctgtcgcaccaaatcagcccctgtgaaactcccacctgtggcctagagtggggtagt
caccccaaagcagcttgctcccccgcctagctcctccaagatctgaagagggtggcttcatctccaggctgcacaagcccgcatttcttcagatgacacgctctaggat
cctgtccatccaggctcgtgctctagcagagcaccacctcctctagaactttctggaagtcttattctaatgagaaggtgatagtcctgtgttctgtttgttggctgctttact
ggttggttaattcattctttcactatctatgcagtgactaccaggtgctgggcagttacgccccttaatggaaatctcagggaagaaatatcctcctctgcctgtgttgatagc
tctattatggcctcaccaaagtgagttatgtgcacgtatgcctctcaccactgaggagatacctccaccactgaggagatacctccagcactgcctatccaagggcaca
gatcatgccacccatgtcactgtcatgagaggcaggcccctgctcttgaagaccatgacttcgttgggacaggtctcatgtcttccttacatctgatgacaactcctactg
cttgataaacaggtgctcaagaagggtttgtagcaggagactaaatcatgaataatacaaggagtccctgggctgaggtgagttagctaagctgggaagtggggaac
tgaggggtgggtcccaggggacttgaaaggcgtcattccagcgaagcccgaaggactcaccatagggacatagtTCTCTTCGCTGTCTCCTGA
GTCTGTGCTGGTGATGCTCTGGGTGGAGATGGGGCGGCAGTACTGGGAGGAGCTGGAGTTGA
e8II AGGTGTGGctgttgtgggaaggaagagttaacactggggaagctgcatggctgctgccatgtttctagcctccagctgggccctgaccctcacacacccaga
cctcagccattagagctgcacactccatcagattaagatctgatatagagcaggtataattattctcttttacagctgaggaaaccaagccacaagacatcgagacatgat
cagctaatgatagcgttagcatcagaatccagctgtcccggcccacaggccagctaccacatcactcattttccctaatggcccgactccaagctcccaccccccacac
atacgcacttacttggccctagaccaagacttggtgataggtgacttgaaggggagttcatcgatgacggtgttgttcctcaggtcaagtggtgttggctttgctggagc
aggaaaagaaagtctggtaatcagccagtaatgataatgctacacatacacaccttatatataacacatgagcatgtttataaaggcctgcaaagtcattccagtttcaca
cagggaaactgacgctcagcgaggttgtataatctttgcccatattgatgagccttgaagtcatgtcttcagattccaaattatatattctttctgctgtatctatggtaattata
ttttcagaatcaaaagtggggacacagggtttgacaacagatttttctccctcaatttatgacttcatttaaaaataaaagactatatataaaatatatttatatattaatatatta
atatttataaaaatatatactatttatatattaacacacatatatgtatttttttgagacggggtctcactctgttgctcaggctggagtgcagcggcgcaatctcggctcactgc
aacctctgcctcctgagttcaagtgattctcctgcctcagcctccctagtagctgggactataggcgcctgccaccacgcttagctaatttttatatttttagtagagacggg
gtttcactattttggccaggctggtctcgaacttctgacctcaggtgatccactcgcctcagcctcccaaagtgttggggttacaggcgtacaggcgtgagccatggcg
cccagccaagagaccatatttgaagtatagatggtcaccgtgactactccctgctgctcctggccaggtcggtgcaagccacagtggaggaaggtagatgccataag
ggcttgagggcttcagttccaaatgctaagatttgaattccctgttggccagacccacaggggggctatctctgcctagctatcctactcctccctttgaatccttgccactg
atccaatgccaggagagtgagattgtagcagacgaggaccttgcctggtctacatgcaggggggcacaaagctgactttactgaacgcctgctatgtaccaaagtacta
tcctagatgctctcctttgcagcaaacctgggaggtaggaattaccatctgcatttcacagataagaaaactgagactcagattttacatgatttgccccaagtcacacctc
tcaagtccaggatccacatgaccccctaagcaaatggaaagagtccctagccactgtccagagatgggacaggggaaagaatggacttacctttccgatcaggtttga
ggttgcggttgacaggggggtggctgaatctcacttcctctgctggggcctcggtgcacaggaagggttgttgatgggtagccaagtgagtcaaagtgatgggcccct
gggctcattgggatgtagacgctctgggaattatcacctgctcgttccatggccaacagggtggaagaacctggattcatgggcacatagttgtcttcagaattggtgct
gtccgatcggcccacaatcattttccctggctagggagaggaacagtgaaagaaatacagctgttactagcaagaagaaacaggaaggggggtaagactttgtaaatg
aggcaagacaagtcatgactgggttagctatactttgaagctgtaagggagaccttggggagaccaaataagaaagcagagtagtgggaggctgaggcgggtgga
tcacttgaggtgaggggttcgagaccagcttggccaacatggcaaaacctcctctctactgaaaatccaaagtattagctgggcgtggtggcatatgcttgtagtccca
gctactcaggaggctgaggcacgagaatcgcttggagctgggaggccaaggttgtagtgagccgagatcatgccactgcactccagcctgggtgacagagtgaga
ctctgtttcaaaaaataaaaaataaaaaaaaaggaagcagagtgggagacaagaatatatgaggtctgtgggccccttcaagccttcaggatcctttggcatccactaat
aatttttgataaatctagagtataaaatattcttggtgactgggtccacttggcTTTCTGGACGCCTGACCAACACTGTGGGTGGCAT
e7 AGTCCACAGACTGTAGGTTTGGTTCAGGACCTGCCCTCCACCTCAAGTCCAGACTCACCACTG
GCAGGTCAGCATCTCAGGCCCTGgatcagaacaaggagcatattaggagtcctggctctggaaggtgaggtctggggccagttgtgtttcc
ccagtccaggtttctaagtacctggggctcactccaggcacctaggggaccaccaagctgcagtcagagcttctacagagttgtgtgatacagcaatcccagccttcct
ctaactcttgctctatctgccgtgacatacaactaagctgcttctgccttggttacctgatttcactgtcttttc...3'

Figure 3-18: FISH analysis of m7B. **A:** m7B (labeled with rhodamine) has been hybridized to mouse metaphase chromosomes. Chromosome 16 is paint labeled with FITC. **B:** m7B labeled with FITC shows signal on chromosome 16 and a pair of signals on another chromosome, either chromosome 9 or 10.


A
chr. 16 paint
chr. 16 paint

B
chr 16

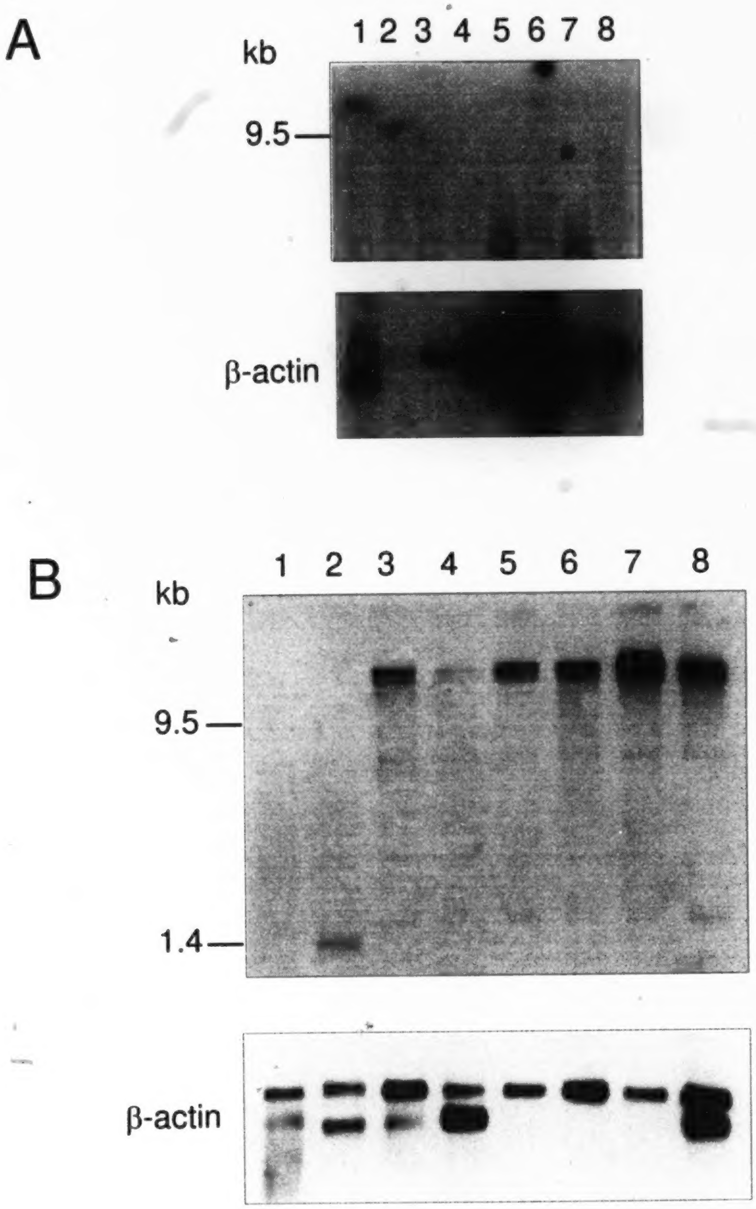
chromosome 22-only cosmid library with YAC 829E11. CES-TK was more accurately mapped within the CESCR by Southern hybridization to YACs, BACs, and PACs and shown to map to the region around D22S9. This putative gene fragment was sequenced by the sequencing facility in the Department of Biological Sciences using an automated sequencing method. A sequence search of this putative exon using BLASTN revealed a significant DNA identity (p-value of 10^{-26}) with the interloukin-9 receptor (IL9R) pseudogene at 10pter, IL9R gene at Xq28 (p-value of 10^{-20}), and IL9R pseudogene at 16pter (p-value of 10^{-20}). The similarity to IL9R mRNA is within the last exon of the gene which is the cytoplasmic domain of the encoded protein. Although IL9R seem to lack a catalytic domain, it is involved in tyrosine phosphorylation by binding to a kinase known as Jak (Ihle and Kerr, 1995). The CES-TK sequence contains a 165 bp ORF in the +1 frame, however, the sequence of this fragment has not yet been confirmed with manual sequencing. The sequence of CES-TK is provided in the appendix. Southern hybridization of this putative gene fragment to a somatic cell hybrid panel demonstrated hybridization to only chromosome 22 at high stringency but when the stringency was lowered many bands were detected (not shown). Northern hybridization detected a large transcript (>10 kb) in the adult brain and heart (figure 3-19). cDNA libraries made from fetal brain and adult heart were screened, with no success in obtaining any cDNA. However, screening a mouse 8.5 dpc library (provided kindly by Dr. Rancourt) led to the isolation of a 0.9 kb cDNA (mtk). Northern hybridization of this cDNA to RNA from different adult and fetal mouse tissues detected a large (>10 kb) transcript (figure 3-19).

2. *PAX6, pnut, VEGF*

I have also attempted to search in the CESCR for other homologous sequences to previously identified and characterized genes. The candidate genes were selected either based on their developmental function (and therefore were suspected to have a part in the features of CES) or on the basis of their mapping to the proximal chromosome 22.

The primers for the three exons (6, 9, and 12) of the *PAX6* gene representing the homeo domain, paired domain, and ETS domain respectively (Martha et al. 1994) were kindly provided by Dr. M. Walter and were used to amplify these exons from the genomic DNA. The PCR products were hybridized to the YACs (as no BACs or PACs in the CESCR were available at the time) and a somatic cell hybrid panel, at low stringency conditions. However, no significant hybridization indicating the presence of a homologue was observed.

Figure 3-19: Human and mouse Northern analysis of CES-TK and mtk. Panel A shows the hybridization of CES-TK to human RNAs from adult tissues: heart (1), brain (2), placenta (3), lung (4), liver (5), skeletal muscle (6), kidney (7), and pancreas (8). Large transcripts (>10 Kb) are detected in brain, heart, and liver. Panel B shows hybridization of mTK to mouse RNAs from tissues: testis (1), lung (2), ovary (3), heart (4), kidney (5), brain (6), liver (7), and E18 (8). Similarly a large transcript is also detected in all tissues except testis and lung.


A
1 2 3 4 5 6 7 8
kb
9.5
β-actin
B
1 2 3 4 5 6 7 8
kb
9.5
1.4
β-actin

I also investigated the possible localization to the CESCR of a human EST which is similar to the *Drosophila melanogaster* gene *peanut (pnut). pnut* is required for cytokinesis and genetically interacts with *seven in absentia*, a gene involved in neuronal fate determination in the compound eye and which has been suggested to have pleiotropic functions (reviewed in Neufeld and Rubin, 1994). A similar (p-value of 10^{-28}) human sequence (EST cDNA 46484) identified from dbEST has been determined by FISH to map to the proximal region of chromosome 22 (Banfi et al., 1996). This cDNA, EST 46484, was purchased from ATCC and used as a probe on a chromosome 22-only cosmid library. Cosmid 110G9 was isolated and mapped by FISH to the metaphase chromosomes of CM15 (a CES patient with a supernumerary ring chromosome encompassing the CESCR). The FISH experiment was done by D. Shkolny. The result indicated that this cDNA does not localize to the ring chromosome and therefore is outside the minimal CESCR.

The possibility of homologous sequences to VEGF (vascular endothelial growth factor) in CES was also examined by hybridization of a 0.3 kb fragment amplified by PCR to YACs in the CESCR. No hybridization indicating a homologous sequence was detected. This gene has recently been mapped to 6p12-p21 and functions in the development of the vasculature by binding to its receptors such as KDR (Terman et al., 1992).

Part C: Molecular localization of translocations in proximal 22q11

This section explains my studies aimed at defining the translocation breakpoint in a number of patients who were carriers of either *de novo* or congenital translocations involving 22q11. This was originally done to investigate the molecular cause of the high frequency of translocations in the proximal region of chromosome 22. Table 2-3 shows the cell lines of the patients which were used in this study and patient's phenotype. Table 3-2 summarizes the molecular localization of the translocation breakpoint in these individuals. For this purpose, we used FISH of metaphase chromosomes with either phage or cosmids in the vicinity of loci indicated in table 3-2. Below are the characterizations of two cell lines, GM01700 (TOH) and GM07511A, which were studied in further detail.

Translocation mapping of TOH (GM01700)

The initial cytogenetic studies on fibroblasts GM01700 from the patient TOH revealed a balanced translocation involving a breakpoint at 22q11.2: t (21;22)(p12;q11.2) (Cannizzaro et al. 1985). The phenotype of TOH at the age of 5.5 years was upslanting

Locus (clone) / Cell line	D22S458 (N38F3)	D22S9 (107D6)	D22S795 (N63)	D22S789 (N44)	D22S75 (N25)	F6 (C6)	F4 (F4-1)
GM07511A	▨	+	▨	▨	+	+	-
GM03197	▨	+	+	+	–	▨	▨
GM06943	+	–	▨	▨	–	▨	▨
TOH	▨	+	▨	▨	+/-	▨	▨
AT	▨	+	▨	▨	+	▨	▨
BE	▨	▨	▨	▨	▨	▨	–
MS	+	+	+	+	+	▨	▨

Table 3-2: Translocation breakpoint mapping at chromosome 22q11.
The loci are listed from proximal to distal. The FISH probes were prepared by biotin-labeling a clone in the vicinity of each locus. Shaded areas represent the probes which were not used for each cell line or where the result was inconclusive. "+" indicates the probe remained on chromosome 22 and the der(22) of the translocation, indicating that the breakpoint on chromosome 22 is located distal to the locus.

"–" indicates transferring of the probe to the other translocation derivative indicating that the breakpoint on chromosome 22 is located proximal to the locus.

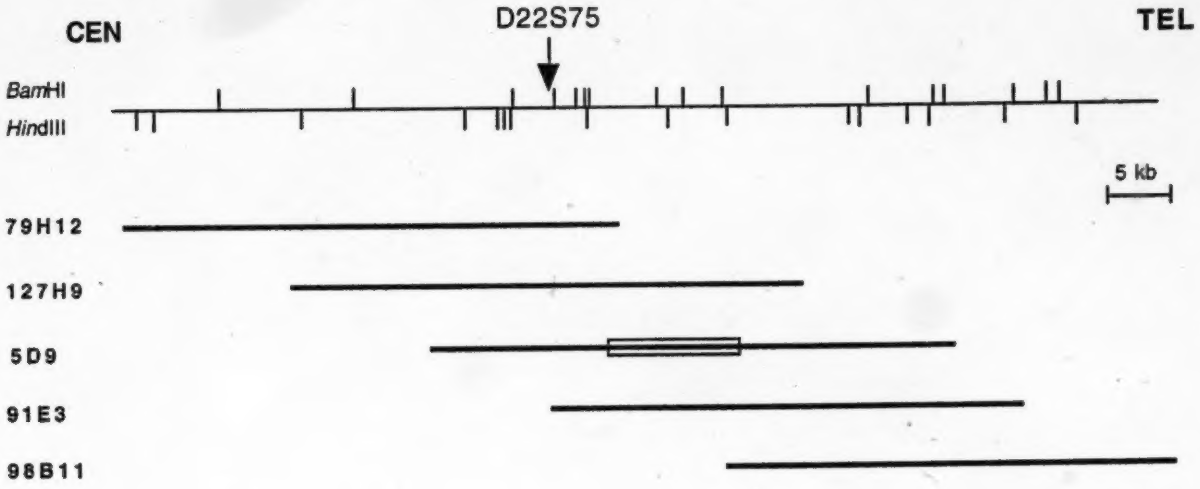
palpebral fissures, epicanthal folds, bulbous nasal tip, and abnormally shaped ears. He was also noted to have hypermobile joints of the fingers and at the wrists. In order to delineate the breakpoint, FISH analysis was done using biotin-labeled cosmids or phages mapping to 22q11.2, together with a control probe which maps to the distal long arm of chromosome 22 (D22S39). Preliminary results localized the breakpoint distal to D22S43, a locus in the cat eye syndrome region, and proximal to the Igλ gene cluster (data not shown).

Further FISH studies mapped the breakpoint around D22S75. The locus D22S75 maps to the CATCH22 critical region and is deleted in the majority of these patients. Reevaluation of the patient phenotype suggested some features of VCFS, a diagnosis not previously given. The patient was reexamined at age 23 by Drs B. Emanuel, and E. Zackai, confirming the presence of features of VCFS.

A cosmid contig around the locus D22S75 was provided by Dr. M. Budarf (Children's Hospital of Philadelphia) (figure 3-20). These cosmids were used in FISH experiments to further delineate the breakpoint. Locus D22S75 was located in cosmid 5D9. Signals on both of the derivative chromosomes (figure 3-21), with approximately equal intensity, were seen in 82% of the metaphase spreads when 5D9 was biotin-labeled and hybridized to the chromosomes of TOH. In 12% of the spreads 5D9 showed a signal only on the der(22) and in 3%, the signal was seen only on the der(21) (figure 3-21). To narrow the translocation breakpoint further, we used cosmids on either side of 5D9 (figure 3-21). FISH analysis with the next proximal cosmid (127H9), which has approximately 20 kb overlap with 5D9 (figure 3-21), showed hybridization on both of the derivatives in a minority of cells. The der(22) signal was stronger and present in most of the metaphase spreads while the der(21) signal was weaker and absent in the majority of spreads (figure 3-20). The most proximal cosmid in the contig (79H12) which overlaps by approximately 13 kb with 5D9, demonstrated no hybridization to der(21), but strong signal on der(22) (figure 3-21). This indicates that the TOH translocation breakpoint is located distal to 79H12 and not within the proximal part of 5D9.

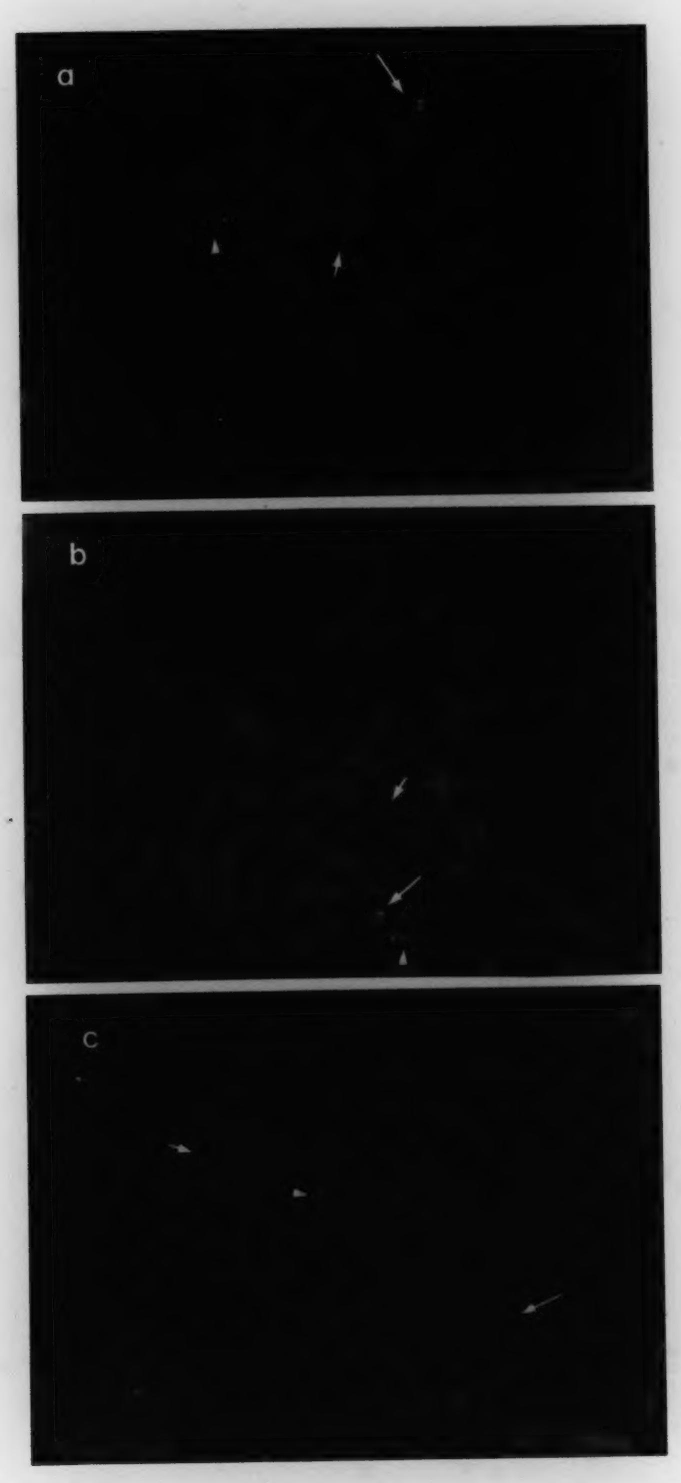
The most distal cosmid, 98B11 , which overlaps 5D9 by approximately 18 kb (figure 3-21), showed no hybridization to the der(22) in any of the metaphase spreads examined, but strong signal on der(21) (figure 3-20). This indicates that the TOH breakpoint is located proximal to 98B11 and excludes the majority of the overlapping region of 5D9. Taken together, these results indicate that the TOH translocation breakpoint can be localized to approximately 10 kb in the middle of 5D9, between the region covered by cosmids 79H12 and 98B11. Our FISH results could not exclude the fact that the

Figure 3-20: Cosmid contig adjacent to D22S75 and FISH analysis of these cosmids in a patient (TOH) carrying a t(21;22)(p12;q11.2). The top section shows a restriction map of the region around D22S75. Below the map are shown a number of cosmids adjacent to D22S75 which were used in this study. The open box indicates the most likely location of the translocation breakpoint as determined from our FISH analysis. Numbers in the table represent the frequency of hybridization to the TOH normal chromosome 22 and the translocation derivatives. From 25 to 30 spreads were examined in each case. The metaphase spreads which did not show a clear hybridization of either the control probe (D22S39) or the test probe were excluded from the study. Restriction map and cosmids were provided by Dr. M. Budarf.


CEN
D22S75
TEL
BamHI
HindIII
5 kb
79H12
127H9
5D9
91E3
98B11

cosmid probe	der(21) and der(22)	only der(21)	only der(22)
79H12	0	0	90
127H9	34	1	61
5D9	82	3	12
91E3	40	53	0
98B11	0	95	0

Figure 3-21: FISH analysis of the cosmids in the D22S75 region hybridized to metaphase spreads from TOH. Panels a, b, and c demonstrate hybridization with the biotin labeled cosmids 79H12, 5D9, and 98B11 respectively. A cosmid probe for D22S39, which hybridizes to distal 22q13.3 was used to mark chromosome 22. The chromosome 22, der(22), and der(21) are indicated by a long arrow, short arrow, and arrowhead respectively. 5D9 shows hybridization to both der(22) and der(21), indicating that the breakpoint lies within this cosmid (b).


a
b
c

breakpoint could map to the most distal region of 79H12 or the very proximal region of 98B11.

Further molecular analysis was done by Susan Holmes (Children's Hospital of Philadelphia). This project was transferred to Dr. B. Emanuel's laboratory due to a long standing arrangement between our laboratories concerning the study of CES and the 22q11 deletion syndromes.

Translocation mapping in GM07511A:

This patient carries a balanced translocation t(14;22)(q32;q11.2) and has been diagnosed with the Gardner syndrome (MIM# 175100), which is considered unrelated to the translocation. FISH analysis originally located the breakpoint distal to the phage clone G12 (not shown in figure 3-22) and proximal to F4, and therefore within the IGL-V region. Further FISH analysis with the biotin-labeled phage clones, G9 and F6 (figure 3-22) demonstrated the hybridization of these clones to the der(22) of the translocation indicating that the breakpoint localizes distal to the region represented by these clones. The phage clone F4 (figure 3-22) hybridized to the der(14) of the translocation. My preliminary result with the clone F7 indicated hybridization to the der(22) and therefore delineating the breakpoint to the region between F7 and F4, a region of approximately 100 kb.

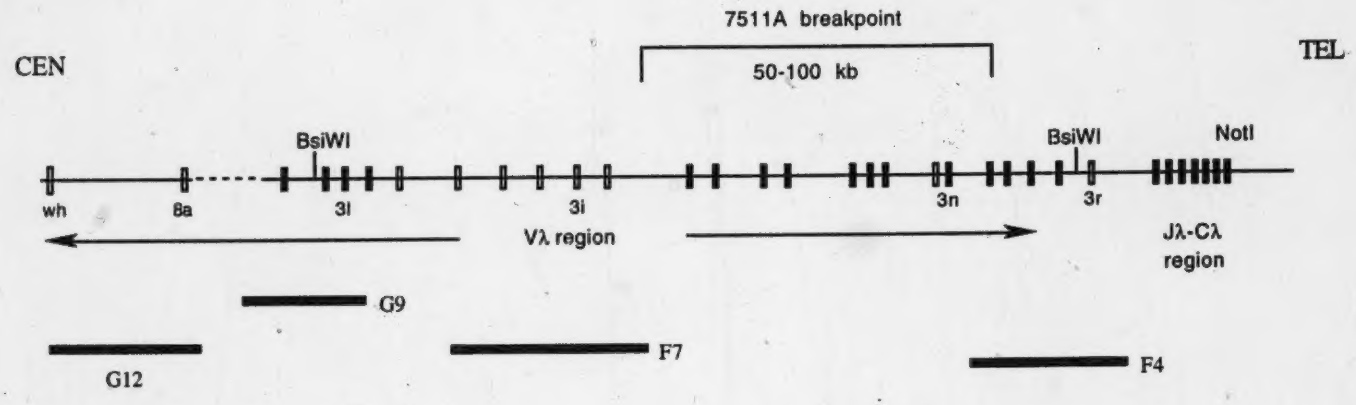

7511A breakpoint
50-100 kb
CEN
TEL
BsiWI
BsiWI
NotI
wh
8a
3l
3i
3n
3r
Vλ region
Jλ-Cλ region
G9
G12
F7
F4

Figure 3-22: A partial map of the IGL locus on chromosome 22. The part of the map represented by clones G12 to F4 is redrawn, roughly to scale, from Frippiat et al. (1995). The solid boxes represent the Jλ-Cλ genes and the open boxes demonstrate the Vλ genes. The translocation breakpoint maps to a 50-100 kb region shown on the figure.

Chapter IV: Discussion

Part A: Cloning within the CESCR

This study was undertaken to clone genes within the cat eye syndrome critical region (CESCR), which encompasses a region of approximately 2 Mb adjacent to the centromere on the q arm of chromosome 22 (McDermid et al., 1996). On the basis of DNA studies of a CES individual with an interstitial duplication (Mears AJ, Ph.D. thesis, 1995), a smaller provisional critical region has been defined. This region extends from locus D22S795 to locus D22S57, spanning approximately 1 Mb (figure 1-3). The duplication of this region is presumed to be responsible for most of the features seen in these patients including preauricular skin tags and pits, heart defects, and renal defects. Two other common features of CES, imperforate anus and coloboma, could still map to this region (discussed in chapter I). In this study I mainly concentrated on this provisional CESCR both for cloning and gene identification attempts.

A contig of the CESCR composed of YACs, BACs, and PACs

YAC contigs which cover approximately half of the CESCR have been produced (McDermid et al. 1996). The loci content of the YACs (obtained from the Chromosome 22 Genome Center) was investigated by Southern analysis using 14 probes in the region. As well, the size of the inserts was determined by PFGE (McDermid et al. 1996). Many of the YACs used in this study were not shown in figure 3-2 in the previous chapter, due to the finding that the majority of these YACs were partially deleted or hybridized to other chromosomes by FISH analysis (McDermid et al. 1996, McTaggart K, MSc thesis, 1997). One explanation for the hybridization to other chromosomes as well as chromosome 22 would be chimerism. This was proven for YAC 734B10 by mapping end clones, and suggested for others by comparison to the PFGE map. An alternative explanation for the high degree of apparent instability of the YACs in this region could be crosshybridization due to the presence of pseudogenes and of repeats which have homologues elsewhere in the genome (discussed below). Both phenomenon are probably involved. In any case, despite the advantage of carrying large inserts, YACs are unstable and prone to deletion and other rearrangements as demonstrated in this study and by others (reviewed in Kim et al., 1996). This has been speculated to be the result of recombinations between repetitive sequences present in a large fragment of human DNA. Recently a contig of bacterial artificial chromosomes (BACs) and P1 artificial chromosomes (PACs) encompassing the 1 Mb provisional CESCR has been produced (A. Johnson et al., unpublished results). BACs and PACs are ideal for mapping due to their stability and their size, which is amenable to direct sequencing. The relative stability of BACs in comparison with YACs partly relates to the bacterial system which is used for the propagation of these vectors, where a recombination-deficient strain can be utilized for construction of a BAC library leading to a higher stability (Shizuya et al., 1992). BACs stay in low copy numbers (1-2 copies) in bacterial cells, leading to a lower probability of recombination between the BACs. The BACs and PACs used to clone the CESCR were isolated by hybridization to the unique loci in this region. The gaps in the contig after the initial assembly of the isolated clones were filled by end-cloning and rescreening a BAC or PAC library. The BAC/PAC contig produced in this region should be of great value in the identification and characterization of the genes in this region.

Part B: Identification of genes in the CESCR

The main purpose of this study was to identify genes located in the critical region for cat eye syndrome. Studies in the gene-rich segments of the human genome have suggested a gene density of one gene per 10-50 kb (Ansari-Lari et al., 1997), consequently, the 1 Mb CESCR may contain a minimum of 20 genes. One plausible explanation for the etiology of the abnormal features in CES is the overexpression of a gene(s) located in this region with a function in the development of organs such as eye, heart, and kidney. Therefore characterization of the CESCR could identify a gene(s) important in the development of these organs. One intriguing question is how many genes are involved in CES? Since a number of malformations with apparently no similar developmental origins are involved, one would assume that such a syndrome is the result of defect in the expression of more than one gene. Studies of other syndromes such as Alagille (MIM# 118450), Angelman (MIM# 105830), and CATCH22, have helped in our understanding of conditions caused by deletion or duplication of large segments of the genome (segmental aneusomy). Alagille syndrome patients have cardiac, ocular, and skeletal defects as well as a typical facial appearance. This syndrome is often associated with the deletion of 1.2 to 1.3 Mb in 20p12. Despite the various developmental abnormalities, a single gene*JAG1* (which is a ligand for the Notch transmembrane receptor) has been identified from the critical region and shown to be mutated in four families with no apparent deletion of the critical region (Li et al., 1997). Similarly, the recent findings of mutations in *UBE3A* (a ubiquitin protein ligase) which lies in the critical deletion region of Angelman syndrome at 15q11-13, has suggested that in some segmental aneusomy conditions, the disrupted expression of one gene might be responsible for the various developmental defects (reviewed in Budarf and Emanuel, 1997). On the other hand, Williams syndrome phenotype (which includes heart defects, growth and mental retardation, connective tissue abnormalities, and characteristic abnormal facial features) is associated with deletions of 7q11.23 and likely caused by haploinsufficiency of at least three genes (Osborne et al., 1996). For CATCH22 it is still not clear how many genes are involved. However, the finding of two balanced translocations, ADU and TOH (this study), each producing a number of features of CATCH22, supports the hypothesis that a full phenotype of CATCH22 is caused by deletion of more than one gene (reviewed in Budarf and Emanuel, 1997). The number of genes involved in CES phenotype awaits characterization of the duplicated region.

The only gene identified in this region before this study was ATP6E. This gene is localized to the most distal position of the critical region and is present in 4 copies in patient CM15 whose duplication breakpoint determines the distal boundary of the critical region (Mears et al., 1995). ATP6E, with a cDNA size of 1.3 kb, encodes the E subunit of the vacuolar H^+-ATPase which appears to be expressed ubiquitously (Baud et al., 1994). H^+-ATPase is the main energizer in the vacuolar system of eukaryotes and functions as an ATP-dependent proton pump to generate protonmotive force (Nelson et al., 1991). Although the effect of the overexpression of ATP6E on the features of CES can not be excluded, the widespread expression of this gene does not seem to be in agreement with a tissue-specific dosage-sensitive gene(s) which we speculate to have a part in CES phenotype. Therefore, this study was done to isolate and characterize more candidate genes from the region.

Amplification of putative gene segments

The cosmids, BACs, and PACs were subjected to exon trapping and the isolated exons were further analyzed by sequencing, hybridization to cDNA libraries, Southern and Northern blots, RACE, and computer analysis of sequence. The cosmids used in the exon trapping study were isolated by hybridization of either a unique probe in the CESCR or hybridization of YAC 829D11 to the chromosome 22-only cosmid library as indicated in table 3-1. Overall the exon amplification was successful in identifying putative gene fragments as described in chapter III. All the isolated exons mapped to the provisional CESCR except e1 which is located adjacent to D22S795. The number of exons isolated from cosmids varied from zero to two. A relatively large number of exons (8 exons) were trapped from p238M15 (which is adjacent to D22S43), while no exon was amplified from p120N18 (adjacent to D22S601). This may indicate the difference in the coding potential of the clones located in the distal and proximal positions of the CESCR. Exon trapping has been successfully applied in the isolation of exons from chromosome 22 by other groups (for instance, Troffatter et al., 1995).

An alternative method would have been to apply a "direct cDNA selection" approach for the isolation of genes in this region. This technique involves capturing cDNAs from different cDNA libraries or tissues by hybridization to the DNA from the region of interest. However, I decided to use an expression-independent approach such as exon trapping in order to be able to clone genes with temporal or spatial specificity, or genes with low expression levels.

The identified exons were used to screen a number of cDNA libraries. Several attempts to obtain cDNAs for some of the exons such as e6, e7, e8, and CES-TK failed. For example cDNA libraries from adult heart, fetal brain and tumor cell line Caco-2 were screened for the putative exon e7 with no success in isolating any correct cDNAs. In addition, 3' RACE to amplify the 3' end of the putative gene for e7 was unsuccessful. Overall, a number of genes and gene segments were identified from the CESCR as well as a novel gene KCNMBL which maps to 3q26.3-3q27, the region duplicated in the dup (3q) syndrome.

Identification of a novel gene, KCNMBL, on chromosome 3 and a related sequence from the CESCR

KCNMBLP was isolated from the vicinity of the locus D22S795, a region just proximal to the provisional CESCR. Southern blotting analysis as well as PCR using the primers within this apparently truncated pseudogene demonstrated a second copy on chromosome 3. Using primers located in the exons 2 and 3 of these related sequences, PCR produced a product of similar size from both chromosomes. On this basis we believe the genomic structure of KCNMBL on chromosome 3, at least in the region of exons 2 and 3 should resemble that of KCNMBLP on chromosome 22. Northern hybridization with the 3' RACE product detected a 0.8 kb transcript and a fainter smaller transcript which are expressed relatively equally in all the tissues. The 3' RACE (exon 3) together with cDNA 683888 represent a gene that we called KCNMBL based on the significant homology to a putative regulatory subunit of potassium channel (CO6) in quail and to the human β subunit of Ca^{+2}-activated potassium channel (hslo-beta or KCNMB). Southern hybridization and

PCR analysis of exon 1 indicated the localization of this gene only to chromosome 3. We have presented evidence that the gene on chromosome 3 is the functional copy and that the sequences we cloned from chromosome 22, KCNMBLP, is probably a truncated unprocessed pseudogene based on the following observations:

1. KCNMBLP seems to lack a 5' exon. As shown with Southern hybridization, the 5' exon of KCNMBL maps only to chromosome 3 and not to chromosome 22. This result was confirmed with PCR of chromosome 22 and 3-containing hybrid cell lines.

2. DNA sequence of the cDNA 683888 demonstrates 2.6% divergence over the exons 2 and 3 with the chromosome 22 genomic fragment containing the KCNMBLP, indicating that this cDNA represents a different gene. Comparison of the 3' RACE-amplified exon 3 sequence to the chromosome 22 genomic sequence revealed 9 different nucleotides in 515 bp or 1.8% divergence, while it more closely resembles cDNA 683888 (three different nucleotides). This suggests that the amplified fragment is different from the gene segment on chromosome 22 and probably represents the same gene as cDNA 683888. The apparent lack of amplification of the chromosome 22 copy in our 3' RACE effort supports the hypothesis that this gene segment is not expressed. Alternatively, since only a single RACE product was sequenced, the different nucleotides may be due to mutations randomly produced by Taq DNA polymerase during the PCR amplification. The error rate caused by Taq DNA polymerase could be as high as 1×10^{-3} error/bp under the conditions (1.5 mM Mg^{+2}, pH 8) used in this study (McPherson et al. 1991). Using the equation $f=np/2$, where f is the frequency of mutations, n is the number of amplification cycles, and p is the error rate, the final mutation frequency in my RACE experiment is estimated to be approximately 1 error in every 100 bases in the population of the final amplified DNAs.

3. A 0.8 kb band was detected when the human Northern blots were probed with the 3' RACE product. A very highly similar putative exon exists in KCNMBLP. To demonstrate that the transcript detected represents KCNMBL and not KCNMBLP, I performed RT-PCR using primers within exons 1 and 2, and showed that these exons belong to the same transcript. Hybridization of a probe made from the amplification of exon 1 using P1 and P2 primers did not detect any transcript. I believe this could be due to the small size (~ 100 bp) of the probe.

4. KCNMBLP maps to a region of chromosome 22 which contains an unusually high frequency of repetitive sequences (Minoshima et al., unpublished results) as well as a number of previously identified unprocessed pseudogenes.

Consequently the gene segment (KCNMBLP) on chromosome 22 is most likely a 5' truncated pseudogene of KCNMBL which is not expressed. The fact that KCNMBLP exons 2 and 3 are separated by an intron indicates that this is an unprocessed pseudogene. During evolution a fragment of chromosome 3 encompassing this gene may have been duplicated and then transposed to chromosome 22. The comparison of the sequence of exons 2 and 3 of KCNMBL and KCNMBLP revealed a total of 2.6% divergence, indicating that such an event occurred relatively recently. Based on the estimates of the rates of neutral nucleotide substitution in primates which is suggested to be 1.5x 10^{-9} events per site per year (reviewed in Regnier et al., 1997), I calculate that this duplication and transposition event must have occurred about 17 million years ago which is in agreement with my preliminary data suggesting only one copy in the rhesus monkey (figure 3-6). The rhesus monkey ancestor diverged from the ancestral human 31 million years ago (Takahata and Satta, 1997), presumably before this duplication and transposition occurred. The presence of an ORF in this gene segment indicates that a random mutation event has not yet produced a stop codon in this sequence.

We still can not exclude the possibility that KCNMBLP accounts for the faint ~0.6 kb band detected in our Northern analysis. The apparent lack of amplification of KCNMBLP in 3' RACE could simply be due to the absence of the transcript of this gene segment in the RNA source (9 week fetus) that I used for this amplification. Therefore RNA extracted from other tissues could be examined to confirm this result.

KCNMBL, a putative candidate for some features of dup (3q) syndrome

KCNMBL was further localized to 3q26.3-27 in collaboration with Dr. S. Nayler (University of Texas Health Science Center) using a radiation hybrid mapping panel. This is the minimal region duplicated in the dup (3q) syndrome. The main features of these patients include a characteristic facial appearance, congenital heart anomalies, and mental and growth retardation (reviewed in Rizzu and Baldini, 1994). In addition to mental retardation, this syndrome is also manifested by muscle convulsions (MIM # 122470). I propose that this gene could play a role in the mental retardation and muscle convulsions seen in these patients. Multiple forms of potassium channels have been found to be expressed in various cell types. Potassium channels are involved in producing action potential waveforms in all excitable cells, setting the resting membrane potentials of most cells, and functioning in cellular mechanisms of learning and memory (Salkoff et al., 1994). The voltage-gated potassium channels are the best known group of these proteins with many known subfamilies: *Shaker, Shal, Shab, and Shaw* are conserved in flies, mice and humans. The presence of multiple subfamilies and many members in each subfamily is believed to be caused by gene duplication as well as extensive alternative splicing associated with these genes. Similarly there are many subtypes of Ca^{+2}-activated potassium channels. These channels are regulated by both voltage and internal calcium concentration and are very effective regulators of cell excitability, and indirectly regulate calcium entry (Dworetzky et al., 1996). Despite the different types of these channels, only the genes encoding the subunits of one channel have been identified. The isolated Ca^{+2}-activated potassium channel is composed of at least two subunits, α, and β. The α subunit is the pore-forming subunit while the β subunit (hslo-beta) has a regulatory function (McManus et al., 1995). The β subunit is a membrane protein with two transmembrane regions. As described in chapter III, KCNMBL protein seems to have two strongly hydrophobic regions which are significantly similar to the transmembrane domains of hslo-beta, as well as the conserved region for N-glycosylation, suggesting that KCNMBL is also a membrane protein with an extracellular domain and that it could have a similar function as a regulator of a subtype of the potassium channels. Since Ca^{+2}-activated potassium channels have essential function in the nervous system and are important in determining the cell excitability, the duplication of KCNMBL and therefore the possible overexpression of it in the dup (3q) syndrome could explain some of the features such as the mental retardation or the muscle convulsion seen in these patients.

Pericentric region of chromosome 22 is rich in unprocessed pseudogenes

The discovery of KCNMBLP has added one more pseudogene to those previously identified in chromosome 22q11 (figure 1-3). This region is known to be rich in low copy number repeats and gene families, including anonymous DNA fragments and expressed sequences (reviewed in Collins et al., 1997). One example is the γ-glutamyl transferase (GGT) gene family with up to 7 copies in chromosome 22q11. Courtay et al. (1994) proposed that one GGT locus encoded a ubiquitous transcript while four GGT loci were transcribed in a tissue-specific manner. One inactive copy (pseudogene) of GGT (GGT12), which is probably a 5' truncated gene segment with a stop codon in position 1930 bp, maps to the CESCR in the region of D22S43 (figure 1-3). There is only one GGT locus in rat (reviewed in Collins et al. 1997) therefore, this duplication has probably occurred after the divergence of humans and rats ancestors. The low copy number repeats are dispersed in 22q11 (such as GGT and D22S131, shown in figure 1-3). This region has been suggested to be unstable, leading to interstitial deletions and duplications in CATCH22 and CES respectively (McTaggart K, Msc thesis, 1997). The instability could be attributed to misalignment and crossing over between these repeats, as has been observed for the STS locus on chromosome X (Yen et al., 1990) or the duplication on 17p causing Charcot-Marie-Tooth disease type 1A (Pentao et al., 1992).

In addition to the low copy repeats, an unexpectedly large number of truncated, unprocessed pseudogenes have been localized to proximal 22q11 in the last two years. Pseudogenes exhibit sequence similarity to functional genes, but they can not be translated into functional proteins because of mutations disrupting transcription, splicing, or translation. Some processed pseudogenes lack introns, usually have poly(A) tail, and are flanked by repetitive elements. These pseudogenes are believed to be generated by reverse transcription and the subsequent insertion into the genome (Maestre et al., 1995). This class of pseudogenes is generally believed to be nonfunctional, however, there are examples of genes with all the characteristics of a processed pseudogene which are expressed (reviewed in Ansari-Lari et al., 1997). One example is phosphoglycerate kinase 2 (PGK-2) which is shown to have the characteristics of a processed pseudogene but has retained its ORF with 85% DNA identity with the PGK-1 coding region on chromosome Xq13. PGK-2 is actively expressed in mammalian spermatogenesis (McCarrey and Thomas, 1987). Unprocessed pseudogenes have intact introns, however they lack an essential 5' or 3' portion of the gene or the exons carry crippling mutations which interrupt the normal translation of the gene to protein (Regnier et al., 1997). The latter group is speculated to be the result of duplication of a gene or a portion of a gene and transposition to a new location where it is inactive (Regnier et al., 1997; Eichler et al., 1997). Many pseudogenes of this type have now been localized to the pericentric region of chromosome 22 including those for von Willebrand factor (vWF), neurofibromatosis 1 (NF1), and adrenoleukodystrophy (ALD). The existence of several related loci for these genes could produce difficulty in the molecular diagnosis of the associated diseases, therefore, the study of these sequences is of a great significance. The vWF gene is located at 12p13.3 and contains 52 exons encompassing approximately 178 kb, which encodes the von Willebrand factor. Mutations in this gene causes a form of hemophilia known as von Willebrand disease (reviewed in Eikenboom et al., 1991). A segment of a closely related sequence to this gene has been isolated from the CESCR in the D22S50 region (figure 1-3) which covers from exon 23 to 34 of this gene, a region of approximately 29 kb (Mancuso et al., 1991). The sequence of this pseudogene has diverged about 3% from the active gene. Eikenboom et al. (1994) studied two families with multiple mutations in the vWF gene and showed that the mutations correspond to the sequence of the pseudogene on 22q11. An interchromosomal gene conversion event between the pseudogene and the functional gene has been proposed to be the underlying cause of these mutations (Eikenboom et al., 1994).

Similar pseudogenes have been found for NF1 on chromosome 17q11.2. The NF1 gene has an estimated *de novo* mutation rate of 1x 10^{-4} which is one of the highest among the human disorders (Huson et al., 1989). Southern and FISH analysis by several groups has demonstrated a number of related sequences to this gene on chromosomes 2, 12, 14, 15, 20, 21, and 22 (reviewed in Hulsebos et al., 1996). Sequence analysis of several fragments of these pseudogenes demonstrated that the chromosome 22 derived NF1 sequence which maps to the proximal region of CESCR (figure 1-3) is the most divergent with 90.6% identity within the sequenced region (exon 13 to exon 15 of the NF1 genomic sequence). This degree of similarity marks the beginning of this divergence to approximately 30 million years ago (Regnier et al., 1997). The pseudogene on 22q11 covers exons 7-28 of NF1, a region of about 100 kb (Regnier et al., 1997) within the CESCR (Hulsebos et al., 1996). The sequence alignment of the pseudogenes with the NF1 sequence showed that the exon-intron organization was well conserved within the pseudogenes. Similar duplication and transposition event also accounts for the multiple related sequences of the Adrenoleukodystrophy gene (ALD) on chromosome bands 2p11, 10p11, 16p11, and 22q11 (Eichler et al., 1997). These sequences are truncated unprocessed pseudogenes resulting from the duplication and transposition of a ~10 Kb segment of the active ALD gene on Xq28 (Eichler et al., 1997). An interesting observation is the centromeric localization of these pseudogenes. Sequencing adjacent to ALD gene segment on 16p11 has revealed satellite III sequence homology which could indicate a role for these repeats in the movement of these so-called duplicons (Eichler et al., 1997).

My finding of another truncated unprocessed pseudogene, KCNMBLP, in 22q11 is further evidence indicating that the proximal regions of at least chromosome 22 contains a high number of these sequences. In addition, a DNA fragment representing another duplication derived from Xq28, originally discovered by Dr. Eichler, maps just proximal to KCNMBLP (Banting, unpublished results). These findings might also help in explaining the lack of success in obtaining cDNAs for a number of our trapped exons in this study. For example several attempts to identify a cDNA for e7, e8, as well as CES-TK (which has been amplified using degenerate primers) failed. These could represent unprocessed pseudogenes which have diverged from the functional copy and as a result do not represent any active gene. The relatively high similarity (~75% DNA identity and 50% protein similarity) of CES-TK with the IL9R tyrosine kinase on Xq28 as well as IL9R pseudogenes on chromosome 10 and 16 may account for another unstable segment of Xq28 which has been transposed to the CESCR much less recently allowing greater divergence. In the light of this emerging concept we could explain some of the high frequency of chimeric YACs we obtained in this region. This chimerism could in fact represent the existence of many DNA segments which have been duplicated and transposed to 22q11. Further sequence analysis in the region would clarify this matter. Eichler et al. (1997) has suggested that such a pericentric plasticity could have some evolutionary significance by juxtaposing different cassettes from diverse genes and creating a reservoir of genes in the genome with potentially new functions.

Since at least four of these pseudogenes cluster at or are centromeric to most proximal region of the CESCR, and functional genes have been found within the region more distally, there may be a gradient, with more repeats and pseudogenes the closer one is to the centromere. Exon trapping was originally chosen because this approach is not expression-dependent, and therefore would isolate rare and temporally-specific transcripts. The presence of pseudogenes in this region indicates that a different approach is needed to isolate genes from the CESCR, since exon trapping does not distinguish between genes and pseudogenes. An expression-based approach such as direct cDNA selection could probably alleviate this problem. However, the CESCR is scheduled to be sequenced by the end of 1998, and therefore, the gene identification can be performed by searching for ESTs (*in silico* cDNA selection) more efficiently.

IDGFL, a putative candidate for some features of CES

Exon trapping in a more distal region of the CESCR was successful in identifying the putative gene IDGFL. The adult expression of IDGFL is more restricted to the lung, while the fetal expression is more evident in liver, lung, and placenta. My Southern analysis of e5 detected a single band which localizes to chromosome 22, suggesting that this putative exon is not a sequence closely related to another gene in the genome, and therefore, probably is not a pseudogene. The cDNA was isolated from the TIGR database by searching with the sequence of the putative exon e5, which is adjacent to D22S43. The cDNA is 3.7 kb long and is similar to the size of the transcript which was detected on the Northern analysis of e5. However, the sequencing of this clone revealed that e5 is at the most 5' end of the clone. Since e5 is the most 5' exon in the identified IDGFL cDNA, it can not be the first exon of this gene otherwise it would not have been trapped by exon trapping, which requires the conserved 5' and 3' splice junctions, and therefore, the 5' end remains to be cloned. The more 5' sequences could be identified by a 5' RACE approach in order to clone the whole gene.

The identified cDNA contains an ORF (ORF1, figure 3-14) coding for 525 amino acid residues. A second large ORF (525 bp, ORF2, figure 3-14) exists in frame with ORF1, suggesting that a sequencing error might have produced a stop codon disrupting a single large ORF. On the other hand, the DNA sequence of ORF2 overlaps with the region similar to repetitive sequences (Alu, and L2), and therefore, is less likely to represent a coding region. The protein encoded by ORF1 is similar in sequence to two previously identified proteins, IDGF, and AGSA. IDGF has been isolated from the embryonic cells of the flesh fly (*Sarcophaga peregrina*) and is presumed to function as a growth factor, since it stimulates the proliferation of embryonic cells of the flesh fly in conditioned medium, with a specific activity comparable to mammalian growth factors. The expression of significant amount of IDGF at various developmental stages (when a high rate of cell proliferation is required) supports the function of IDGFL as a growth factor (Homma et al. 1996). Similarly, AGSA was identified from the dense core vesicles (which store secretory proteins) of the atrial glands of *Aplysia californica*. However, the function of this protein is still unclear (Sossin et al. 1989). The similarity of IDGFL with IDGF and AGSA suggests that it may also function as a growth factor. The abnormal overexpression of a putative growth factor such as IDGFL may interfere with the normal development of some the organs affected in CES patients by increasing cell proliferation. Some features of CES such as preauricular skin tags and imperforate anus could be explained by such a mechanism, since they seem to be the result of abnormal cell proliferation or cell death.

To do developmental expression studies e5 was hybridized to a mouse cDNA library and a cDNA (mC1) was identified. The similarity mC1 to IDGFL was also confirmed by hybridization with a fragment of IDGFL which did not include e5 . The Northern hybridization of this clone detected a transcript of approximately 1.4 kb in size which is ubiquitously expressed. e5 was also hybridized to mouse RNA samples on a Northern blot and demonstrated a faint 1.5 kb as well as a faint 2.5 kb bands (not shown). The 2.5 kb band seems to be present in all tissues examined, while the 1.5 kb transcript expression is restricted to the mouse liver. This was somewhat of a surprise since we expected that the orthologous mouse cDNA to be of approximately the same size of the IDGFL cDNA (~3.7-4 kb). This difference may be attributed to the difference in the length of the untranslated regions (UTRs). The IDGFL cDNA contains a fairly large (2193 bp) 3' UTR which could account for this difference. A comparison of the mRNA and protein sequences of 1196 orthologous mouse and human genes has demonstrated 85% identity on average between the coding regions whereas the 5' and 3' UTRs are less conserved (~67% identity) in sequence (Makalowski et al. 1996). No general comparison of the sizes of the UTRs of the human and mouse transcripts has been reported. However, the lower rate of conservation in sequence of the UTRs may suggest a larger variation among the size of the UTRs in human and mice. Since the size as well as the expression pattern of the transcript detected by mC1 is different, this cDNA may not represent the true ortholog of IDGFL in mice. When e5 was hybridized at low stringency to a Zoo blot, it did not detect any band in mouse (data not shown). However, the mouse DNA used was not of good quality and the experiment needs to be repeated.

A putative gene, SAHL1

The sequence similarity search in dbEST database of e11 and e22 resulted in the identification of SAHL1. SAHL1 seems to be located only on chromosome 22 as suggested by my hybrid panel hybridization to e22 (figure 3-9). The putative exon e22 also hybridizes to DNA from a number of other organisms (figure 3-9) indicating that SAHL1 may be a functional gene. Although the analysis of the sequence of the isolated cDNAs and c1A3 genomic clone suggested the presence of at least three splice variants for SAHL1, Northern hybridization with the SAHL1-1 fragment (which is shared between these cDNAs) revealed two transcripts, 2.1 and 2.8 kb in size, in the liver and kidney. This could be due to specific expression of one of these cDNAs in the tissues not examined in this study. Alternatively, the transcript sizes may be similar, and therefore, not resolved on the Northern blot which was used. In addition to the liver and kidney, SAHL1-1 cDNA (EST 23249) must be expressed in the adipose tissue, since it was originally obtained from this tissue.

Analysis of the ~ 36 kb sequence of cosmid c1A3 containing SAHL1 revealed 7 exons and 3 ESTs. Since the orientation of cDNA 32483 agrees with SAHL1, it could represent the 3' end of the SAHL1-3 cDNA (figure 3-10). This can be investigated by RT-PCR using primers within this cDNA and SAHL1-3. The cDNA 32483 has been isolated fom the fetal brain RNA, indicating that this gene is expressed in the brain. The EST cDNA FB20C7 was demonstrated to detect a different transcript on our Northern blot indicating it is different from SAHL1. The other EST detected by the sequence of c1A3 is 41a1, which is from a randomly primed retinal cDNA library. This cDNA was not available for further studies. An interesting observation was that the sequence of c1A3 was relatively poor in exons (as predicted by Grail) and rich in repetitive sequences (data not shown). This may represent the poor coding potential of the sequences in the proximal regions of the CESCR.

Although the three cDNAs SAHL1-1, SAHL1-2, and SAHL1-3 share exons e11, e22, and exon a (figure 3-11), they have different 3' ends, indicating that this putative gene (SAHL1) undergoes alternative splicing. The 3' exon of SAHL1-2 overlaps with the 3' end of a cDNA (EST 41a1, figure 3-10) which demonstrates a different exon/intron boundary and therefore probably accounts for a fourth splice variant of SAHL1. This cDNA which was isolated by *in silico* cDNA selection to previously identified sequences in database is from a randomly primed retinal cDNA library, suggesting its expression in the eye (one of the organs affected in CES).

The strong conservation of SAHL1-1 in human and mouse is suggested by the hybridization of SAHL1-1 to mouse RNA at high stringency (figure 3-10). No mouse cDNA has been identified yet for SAHL1-1, despite screening an 8.5 dpc cDNA library. Considering the expression of SAHL1 in the kidney and liver, screening of a mouse kidney or liver cDNA library might result in the identification of the cDNA. I have some preliminary evidence of the expression of this gene during embryonic development. The whole mount *in situ* hybridization results suggested hybridization to the developing kidney region in 11.5 dpc embryos. The formation of the permanent kidney in mouse starts in around day 11 by the growth of the uretric bud into metanephric mesoderm (Kaufman 1992). If this gene is confirmed to be expressed in the developing kidney, the overexpression of this gene could have some disrupting effect on the normal development of this organ as seen in CES.

Comparative mapping of the CESCR

In the previous chapter I presented my results in identifying a mouse genomic clone for e7 as well as SAHL1. The identified phage clone for e7 (m7B) which showed the strongest hybridization to this gene segment was mapped by FISH to the mouse chromosomes (figure 3-17). This was done in collaboration with the CGAT Cytogenetic Laboratory Services at the Hospital for Sick Children. This result indicated the hybridization of this clone to mouse chromosome 16 on the border of the bands B and C in 75% of the metaphase spreads. A second less common (30% of spreads) hybridization is observed on either chromosome 9 or 10. This is also in agreement with our finding of two positive genomic clone in mouse which hybridized to e7. This result provides preliminary evidence suggesting that mouse chromosome 16 may contain the homologous region to the CESCR. The cDNA c7/8 which has significant similarity to e7 does not have any ORF, and therefore, could be a pseudogene .

A similar study was undertaken to determine a positive mouse genomic clone for the cDNA SAHL1-1. The screening of a mouse BAC library resulted in a positive clone, which is being mapped by FISH in collaboration with Dr. J. Kornberg (UCLA).

The recent mapping studies by T. Footz (unpublished data) has indicated that ATP6E maps to mouse chromosome 6. This was somewhat of a surprise considering that the more proximal region (e7, as known from my result), and distal region (CATCH22) are homologous to mouse chromosome 16. This can be explained by a rearrangement such as an inversion in human placing the region (ATP6E vicinity) homologous to mouse chromosome 6 in between. Alternatively, the real human e7 gene could be elsewhere in the genome and fortuitously mapping to mouse chromosome 16, while the pseudogene is located in the CESCR.

Developmental defects in CES, a molecular etiology

As explained before, in CES organs such as eye, heart, and kidney develop abnormally. Understanding the molecular etiology of this syndrome requires a detailed knowledge of the embryonic development of these organs at the molecular level. The embryonic choroid fissure normally closes during the seventh week of development, however, when this fails to occur it causes coloboma which is commonly seen in CES. The development of the renal system which is also affected in these patients occurs during week 5 to 8 of the development. The kidney is formed by the growth of the uretric bud into the mass of the metanephric mesoderm. As well the development of the heart starts early and most of the conotruncal malformations, such as TAPVR or TOF in CES patients, are due to the defects occurring in the 3rd to 8th week of embryonic development (Medical Embryology, Sandler, 1985). Therefore we believe the gene(s) whose abnormal expression causes the CES features are very likely to be expressed at the early stages of development. Weleber et al. (1977) suggested the extra chromosome in CES alters the rate of cell division and thus causes the characteristic malformations. As well Walknowska (1977) has suggested that the coloboma, imperforate anus, renal and heart defects in CES can be explained by failure of cellular proliferation and fusion of embryonic clefts, or interaction with other tissue planes or both. Since this syndrome is associated with a duplication of a segment of chromosome 22, the most plausible explanation for the abnormal features would be the overexpression of a dosage sensitive gene(s) located in this region. These genes could be involved directly or indirectly in the development of these organs, and therefore, the overexpression of them somehow affects the normal development of these organs. One such gene could be a growth factor (such as IDGFL identified in this study), a gene which is involved in the pathway of programmed cell death, or a transcription factor which monitors the rate of expression of some genes important in the development of these organs. Since the CESCR contains a high frequency of unprocessed pseudogenes, a possible but less likely hypothesis is that overexpression of these truncated and mutated genes could produce a dominant negative effect on the functional copies of these genes located in other regions of the genome, therefore causing the abnormalities.

An alternative model which I would like to propose is a chromatin modification model. One could assume that the gene(s) whose abnormal expression causes the features of CES are outside of the CESCR on chromosome 22 or on other chromosomes, however, their expression patterns or rates are affected indirectly by the expression of a gene whose product modifies the chromatin state. Consequently the abnormal expression of this gene(s) causes either abnormal repression or activation of some genes which are important in the development of organs affected in CES. Trofatter et al. (1995) have reported the isolation of a chromosome 22 exon (c22-637) which has significant homology to part of a common domain, termed the chromodomain, originally described in *Drosophila*. This protein domain is found in genes whose products associate with heterochromatin such as Polycomb group genes which are known to repress multiple homeotic genes (reviewed in Kennison, 1995). This function by the Polycomb group proteins on homeotic genes is suggested to be through changing the higher-order chromatin structure. Theoretically the putative exon c22-637 or a gene of this type could map to the CESCR and cause the abnormal features when overexpressed. Alternatively, a gene with similarity to Trithorax group genes in *Drosophila* or SNF2/SW12 in yeast could be overexpressed in CES. These genes have been demonstrated to modify chromatin to facilitate the transcriptional activity by gene-specific DNA-binding proteins (reviewed in Kennison, 1995).

Future research

Identification of genes from the CESCR

One finding from this study was that the proximal region of CESCR contains a large number of unprocessed pseudogenes. Consequently, the current gene identification efforts in the CESCR are concentrated on the more distal region of the provisional CESCR. Recent analysis of the genomic sequence of the distal regions of the CESCR has demonstrated a small number of repetitive sequences (McDermid et. al, unpublished result). A number of putative exons which were isolated in this study from the PAC p238M15 are now being characterized in Dr. McDermid's laboratory by P. Brimkmann. The hybridization of these exons to the genome has to be examined first in order to confirm the number of copies in the genome and to indicate that they are not pseudogenes. Subsequently the exons can be used for the isolation of cDNAs either by screening cDNA libraries or by searching sequence for the presence of ESTs. The sequence of the cDNA can be compared with the genomic sequence in order to identify the genomic structure of the genes.

Due to the high frequency of pseudogenes, other approaches for gene identification should be applied to this region. The two more widely used approaches of "cDNA selection" and "identification of genes by a comparative approach" would probably be more productive. The cDNA selection approach would alleviate the problem of trapping unprocessed pseudogenes, however, it has the disadvantage of being limited to the genes which are expressed in certain tissues. Other sequence-based approaches might be utilized when the region is sequenced.

Identification of the region in mouse with conserved synteny to the CESCR

The first step to use a comparative approach to identify the genes in the CESCR is to detect the region in mouse with synteny to the CESCR. In this study we began this by isolating the mouse homologues for e7 and SAHL1-1 gene segments in CESCR. Our preliminary result suggested that a region of mouse chromosome 16 is syntenous in sequence to at least part of the CESCR. This study is being continued for other genes mapping to the distal regions of CESCR by T. Footz in the McDermid laboratory. He has found that ATP6E and the two flanking genes map to mouse chromosome 6 by analyzing a *Mus musculuc/Mus spretus* backcross panel from The Jackson Laboratories. Once the mouse homologues throughout the region have been identified and mapped onto mouse chromosomes, the comparison of the genomic sequence of the mouse region homologous to the CESCR would help us toward identification of more genes from the critical region.

Further characterization of IDGFL, SAHL1, and other candidate genes

Further characterization of these genes might shed a light on the etiology of this syndrome. The full cDNA for these putative genes must be identified first. This can be done by screening other cDNA libraries not used in this study, or by RACE. The temporal expression of these genes should be studied by more detailed expression study at the embryonic and fetal stages in human as well as in mouse. As the genomic sequence of the region will be soon available (Dr. B. Roe, personal communication), the promoter regions of these genes can be identified. Further characterization of expression of these genes then can be carried out by fusing these gene promoters to a reporter gene such as LacZ and studying the embryonic or adult expression in more details. The overexpression of these genes in CES patients must be confirmed by probing Northern blots containing RNA from CES patients. Functional characterization of IDGFL can be performed by purification of the protein and studying the possible function of this gene as a growth factor in cultured cells.

As mentioned in chapter III the human homologue of "*pnut*" maps to proximal 22. This gene is involved in the signaling pathway leading to the development of R7 in the *Drosophila melanogaster* eye as well as being required for cytokinesis. In addition it has been shown to be haploinsufficient in *Drosophila* (therefore dosage sensitive). Although I have demonstrated that the cosmid which is positive for this cDNA (110G9) does not localize to the ring(22) in CM15 which determines the minimal critical region of CES, however, this still could map to the CECs (which are larger than the ring chromosome in CM15) and therefore be overexpressed in most CES patients. Therefore this can be further examined with FISH on the CES patients who carry CECs.

Toward a mouse model for CES

The recent cloning of the CESCR in BAC and PAC vectors (McDermid et al. unpublished results) will allow direct injection of the BACs or PACs carrying different regions of the CESCR into mouse eggs. By studying the phenotype in mice resulting from overexpression of a set of genes on each BAC, the minimal critical region of this syndrome may be narrowed down. Further gene identification study then can be performed in a smaller region, or further characterization of the already identified genes can be limited to the ones mapping to the minimal critical region. Alternatively, the discovery of a region in the mouse with conserved synteny similar to the CESCR would permit mimicking the CES phenotype by duplicating the mouse homologous region. This has been demonstrated to be feasible by inserting loxP sites flanking the desired region and the simultaneous expressing Cre recombinase to stimulate recombination between the loxP sites (Ramirez-Solis et al., 1995). In order to proceed with these experiments, however, it will be necessary to characterize in detail the entire CES region in mice.

Part C: Is there a translocation/rearrangement hot spot on 22q11?

This study was done to investigate the existence of a translocation hot spot in 22q11, and if the translocation breakpoints localize to the common deletion and duplication breakpoints in CATCH22, and CES respectively. Other studies have demonstrated the colocalization of the common deletion breakpoint of CATCH22 with the duplication breakpoint of CECs in a number of CES patients (McTaggart et al., in preparation). As well the distal deletion breakpoint in CATCH22 overlaps with the region involved in the common constitutional translocation between chromosomes 11 and 22 (Budarf et al., unpublished results). My results demonstrated that these translocations did not map to the same location, suggesting that there is no translocation hot spot on this chromosome. Alternatively, there could be a large number of the translocation-prone regions scattered on this chromosome.

A translocation t(21;22)(p12;q11.2) disrupts the CLTCL gene in the CATCH 22 region in TOH

I have presented evidence which demonstrates that the translocation breakpoint in the patient TOH, who has some features of velocardiofacial syndrome, lies within the CATCH22 critical region near locus D22S75. The features seen in patients with VCFS and present in this patient include: upslant of palpebral fissures, epicanthal folds, bulbous nasal tip, abnormally shaped ears, and long slender fingers. This patient does not have a palatal or cardiac anatomic defect. His IQ of 45 is lower than what has been reported for patients with this syndrome (Wang et al., 1996). His diagnosis of cryptorchidism and hypospadias and wide interpupillary distance have been seen in patients with Opitz GBBB who have a 22q11.2 deletion (McDonald-McGinn et al., 1995). Our FISH results demonstrate that cosmid 5D9 in the vicinity of D22S75 hybridizes to both der(22) and der(21) of translocation, indicating the presence of the breakpoint within this cosmid. This could be alternatively explained by the presence of repeats in two different regions separated by the translocation, or the existence of a small duplicated region on both sides of the translocation breakpoint, as is found for the congenital t(11;22) breakpoint. These possibilities were tested by examination of the overlapping proximal and distal cosmids (figure 3-19). The fact that the adjacent cosmids, 79H12 and 98B11, clearly map to one side of the translocation or the other rules out the possibility of repeat elements or a small duplication. FISH analysis with 5D9 overlapping cosmids delineated the translocation breakpoint to a region of approximately 10 kb near the *Not*I site of 5D9 (figure 3-19). The *Not*I site is directly adjacent to D22S75, which has been shown to be deleted in the majority of CATCH 22 patients (Driscoll et al., 1992a,b). The chromosomal rearrangement in TOH was previously reported to be a Robertsonian translocation (Wolff and Schwartz 1992). However, this is clearly ruled out by our results since D22S75 is approximately 4 Mb distal to the centromere (McDermid H, unpublished results).

Further analysis of the breakpoint of TOH has demonstrated that the translocation disrupts the clathrin heavy chain-like gene (*CLTCL*), which is highly expressed in human skeletal muscle, therefore making it a candidate for at least the features of CATCH 22 present in this patient (Holmes and Riazi et al., 1997). The translocation breakpoint resides

5' to exon 28 of this gene and results in a truncated protein lacking the C-terminal 142 residues coded by exons 28-33. Northern analysis of the RNA extracted from the fibroblasts and lymphoblast cell lines of this patient has failed to detect an abnormally sized transcript. This could be due to the instability of the abnormal transcript or a disruption of transcription of the abnormal gene in TOH. Little is known about the function of *CLTCL*. Despite 85% identity to human clathrin heavy chain gene(*CLTC)* at the protein level, the expression pattern of *CLTCL* is very different. *CLTC* is a ubiquitous structural protein of the clathrin-coated pits and vesicles which are involved in endocytosis and membrane receptor trafficking. Clathrin protein forms a lattice on the cytoplasmic face of clathrin-coated pits and vesicles (reviewed in Robinson, 1994). The lattice is formed of overlapping triskelion units made of three heavy and three light chains. Truncation of the 67 residues at the carboxyterminal end of the protein disrupts the trimerization and therefore disruption of the function of the protein (Liu et al., 1995). By analogy the deletion of the C-terminal end of *CLTCL* could cause the same effect if the *CLTCL* gene product forms a triskelion like *CLTC* (Holmes and Riazi, 1997).

A different translocation disrupting the CATCH22 region, t(2;22)(q14;q11.1) in the patient ADU, has also been described (Augusseau et al., 1986). The translocation breakpoint of ADU maps approximately 100 Kb proximal to TOH breakpoint (figure 1-3). The ADU translocation disrupts two novel putative genes DGCR3, and DGCR5. DGCR5 has been demonstrated to be expressed in multiple tissues, however, it does not appear to code for any sizable protein (Sutherland et al., 1995), suggesting that this gene may produce a functional RNA which acts as a modifier of the expression of the surrounding genes. This type of modifier function occurs in other instances such as in XIST, which is solely expressed from the inactive X chromosome and is suggested to produce a functional RNA which prevents the transcription of the majority of the genes on the inactive X chromosome(Brown et al., 1992).

The localization of TOH and ADU translocation breakpoints to different sites in the critical region supports the hypothesis that CATCH22 is a contiguous gene syndrome resulting from monosomy for more than one gene on 22q11.2 (Driscoll et al., 1992a). The term contiguous gene syndrome refers to a group of disorders caused by microdeletion or microduplication of a chromosomal segment encompassing several genes (Schmickel et al., 1986). As well the finding of some but not all features of CATCH22 in TOH supports the idea that this syndrome is a contiguous gene syndrome, and that the *CLTCL* gene disrupted by this translocation may therefore account for only part of the phenotype.

IGL locus involved in translocation t(14;22)(q32;q11.2) in GM07511A

The cell line GM07511 was obtained from a patient diagnosed with Gardner syndrome. Gardner syndrome or familial adenomatous polyposis (MIM # 175100) is characterized by adenomatous polyps of the colon and rectum which can transform into malignant tumors during the patient's life. Another feature of this syndrome is hypertrophy of the retinal pigment epithelium. The gene for this syndrome, APC, maps to 5q21-22 and is suggested to be involved in cell adhesion (Peifer, 1993). The syndrome is therefore thought to be independent of the translocation which involves chromosomes 14 and 22.

My FISH analysis of a lymphoblastoid cell line from this patient indicated that the translocation occurs within the Igλ locus on chromosome 22. The human immunoglobulin lambda (λ) light-chains genes are located in 22q11 and are composed of 20-70 Vλ, and 7-10 Jλ-Cλ genes with the Vλs being proximal to the Jλ and Cλ genes (reviewed in Frippiat et al. 1995). A biotin-labeled phage clone, G12, which represents the most proximal Vλ genes (Frippiat et al., 1995), as well as more distally mapping clones, F6, and F7 were localized to the der(22) of the translocation indicating that the translocation breakpoint maps distal to these clones. FISH with a clone representing the region above the most proximal Cλ gene, however, demonstrated hybridization to the der(14) of the translocation. These results together map the breakpoint to a region of approximately 50-100 kb (figure 3-22). Further FISH analysis with the clones in the region between F7 and F4 is necessary to map the breakpoint. Similar studies needs to be done to localize the breakpoint on chromosome 14. Our preliminary result indicated that the breakpoint on chromosome 14 could be occurring in the immunoglobulin constant gene region (result not shown). This study is of interest because it may represent interchange between similar sequences in the two related loci on chromosome 22 and chromosome 14 which are prone to rearrangements.

References

Aaronson JS, Eckman B, Blevins RA, Borkowski JA, Myerson J, Imran S, Elliston KO (1996) Towards the development of a gene index to the human genome: an assessment of the nature of high-throughput EST sequence data. Genome Res 6:829-845

Abruzzo MA, Erickson RP (1977) Case report, a new syndrome of cleft palate associated with coloboma, hypospadias, deafness, short stature, and radial synostosis. J Med Genet 14:76-80

Allikmets RL, Kashuba VI, Pettersson B, Gizatullin R, Lebedeva T, Kholodnyk ID, Bannikov VM, Petrov N, Zakharyev VM, Winberg G et al. (1994) NotI linking clones as a tool for joining physical and genetic maps of the human genome. Genomics 19:303-309

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403-410

Ansari-Lari MA, Shen Y, Muzny DM, Lee W, Gibbs RA (1997) Large-scale sequencing in human chromosome 12p13: experimental and computational gene structure determination. Genome Res 7:268-280

Antequera F, Bird A (1994) Predicting the total number of human genes. Nature Genet 8:114

Aubry M, Marineau C, Zhang FR, Zahed L, Figelwicz D, Delattre O, Thomas G, De Jong PJ, Julien JP, Rouleau GA (1992) Cloning of six new genes with zinc finger motifs mapping to short and long arms of human acrocentric chromosome 22 (p and q11.2). Genomics 13:641-648

Augusseau S, Jouk S, Jalbert P, Prieur M (1986) DiGeorge syndrome and 22q11 rearrangements. Hum Genet 74:206

Aurias A, Rimbaut C, Buffe D, Zuker J-M, Mazabraud A (1984) Translocation involving chromosome 22 in Ewing's sarcoma. Cancer Genet Cytogenet 12:21-25

Banfi S, Borsani G, Rossi E, Bernard L, Guffanti A, Rubboli F, Marchitiello A, Giglio S, Coluccia E, Zollo M, Zuffardi O, Ballabio A (1996) Identification and mapping of human cDNAs homologous to Drosophila mutant genes through EST database searching. Nature Genet 13:167-174

Baud V, Mears AJ, Lamour V, Scamps C, Duncan AMV, McDermid HE, Lipinski M (1994) The E subunit of vacuolar H^+-ATPase localizes close to the centromere on human chromosome 22. Hum Mol Genet 3:335-339

Bell CJ, Budarf ML, Nieuwenhuijsen BW, Barnoski BL, Buetow KH, Campbell K, Colbert AM, Collins J, Daly M, Desjardins PR, et al. (1995) Integration of physical, breakpoint and genetic maps of chromosome 22. Localization of 587 yeast artificial chromosomes with 238 mapped markers. Hum Mol Genet 4:59-69

Bernardi G, Olofsson B, Filipski J, Zerial M, Salinas J, Cuny G, Meunier-Rotival F (1985) The mosaic genome of warm-blooded vertebrates. Science 228:953-958

Bernardi G (1995) The human genome: Organization and the evolutionary history. Ann Rev Genet 29:445-476

Blennow E, Brondum-Nielsen K (1991) Molecular identification of a small supernumerary marker chromosome by in situ hybridization: diagnosis of an isochromosome 18p with probe L1.84. Clin Genet 39:429-433

Blumenthal T, Spieth J (1996) Gene structure and organization in *Caenorhabditis elegans*. Curr Opin Genet Dev 6:692-698

Blyel S, Ruttenberg HD, Carey JC, Ward K(1994) Familial total anomalous pulmonary venous return: a large Utah-Idaho family. Am J Med Genet 54:462-466

Blyel S, Nelson L, Odelberg SJ, Ruttenberg HD, Otterud B, Leppert M, Ward K (1995) A gene for familial total anomalous pulmonary venous return maps to chromosome 4p13-q12. Am J Hum Genet 56:408-415

Botstein D, White RL, Skolnick MH, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314-331

Britten RJ (1986) Rates of DNA sequence evolution differ between taxonomic groups. Science 231:1393-1398

Brown CJ, Hendrich BD, Rupert JL, Lafreniere RG, Xing Y, Lawrence J, Willard HF (1992) The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. Cell 71:527-542

Brondum-Nielson K, Mikkelsen M (1995) A 10 year survey, 1980-1990, of prenatally diagnosed small supernumerary marker chromosomes, identified by FISH analysis. outcome and follow-up of 14 cases diagnosed in a series of 12699 prenatal samples. Pre Diag 15:615-619

Buckler AJ, Chang DD, Graw SL, Brook JD, Haber DA, Sharp PA, Housman DE (1991) Exon amplification: A strategy to isolate mammalian genes based on RNA splicing. Proc Natl Acad Sci USA 88:4005-4009

Buckton KE, Spowart G, Newton MS, Evans HJ (1985) Forty-four probands with additional marker chromosome. Hum Genet 69:353-370

Budarf ML, Emanuel BS (1997) Progress in the autosomal segmental aneusomy syndromes (SAS)s: single or multi-locus disorders? Hum Mol Genet 6:1657-1665

Budarf ML, Collins J, Gong W, Bruce R, Wong Z, Bailey LC, Sellinger B, Michaud D, Driscoll D, Emanuel BS (1995a) Cloning a balanced translocation associated with DiGeorge syndrome and identification of a disrupted candidate gene. Nature Genetics 10:269-278

Budarf ML, Konkle BA, Ludlow LB, Michaud D, Li M, Yamashiro D, McDonald-McGinn D, Zackai E, Driscoll DA (1995b) Identification of a patient with Bernard-Soulier syndrome and a deletion in the DiGeorge/velo-cardio-facial chromosomal region in 22q11.2. Hum Mol Genet 4:763-766

Budarf ML, Sellinger B, Griffin C, Emanuel BS (1989) Comparative mapping of the constitutional and tumor-associated 11;22 translocation. Am J Hum Genet 45:128-139

Buetow KH, Weber JL, Ludwigsen S, Scherpbier-Heddema T, Duyk GM, Sheffield VC, Wang Z, Murray JC (1994) Integrated human genome-wide maps constructed using the CEPH reference panel. Nature Genet 6:391-393

Buhler EM, Mehes K, Muller H, Stadler GR (1972) Cat-eye-syndrome, a partial trisomy 22. Humangenetik 15:150-162

Bulle F, Mattei MG, Siegert A, Pawlak E, Passage E, Chobert MN, Laperche Y, Guellaen G (1987) Assignment of the human gamma-glutamyl transferase gene to the long arm of chromosome 22. Hum Genet 76:283-286

Burke DT, Carle GF, Olson MV (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. Science 236:806-812

Burn J, Goodship J (1996) Developmental genetics of the heart. Curr Opin Genet Dev 6:322-325

Burn TC, Conners TC, Klinger KW, Landes GM (1995) Increased exon-trapping efficiency through modifications to the pSPL3 splicing vector. Gene 161:183-187

Burset M, Guigo R (1996) Evaluation of gene prediction programs. Genomics 34:353-367

Cannizzaro AL, Aronson MM, Emanuel BS (1985) In situ hybridization and translocation breakpoint mapping II. Two unusual t(21;22) translocations. Cytogenet Cell Genet 39:173-178

Carey AH, Kelly D, Halford S, Wadey R, Wilson D, Goodship J, Burn J et al. (1992) Molecular genetic study of the frequency of monosomy 22q11 in DiGeorge syndrome. Am J Hum Genet 51:964-970

Carey AH, Roach S, Williamson R, Dumanski JP, Nordenskjold M, Collins VP, Rouleau G, Blin N, Jalbert P, Scambler PJ (1990) Localization of 27 cDNA markers to the region of human chromosome 22q11-pter deleted in patients with the DiGeorge syndrome and duplicated in the der22 syndrome. Genomics 7:299-306

Castle WE, Wachter WL (1924) Variations of linkage in rats and mice. Genetics 9:1-12

Chance PF, Alderson MK, Leppig KA, Lensch MW, Matsunami N, Smith B, Swanson PD, Odelberg SJ, Disteche CM, Bird TD (1993) DNA deletion associated with hereditary neuropathy with liability to pressure palsies. Cell 72:143-151

Chandrasekharappa SC, Guru SC, Manickam P, Olufemi SH, Collins FS, Emmert-Buck MR, Debelenko LV, et al. (1997) Positional cloning of the gene for multiple endocrine neoplasia-type1. Science 276:404-406

Chen EY, Zollo M, Mazzarella R, Coccodicola A, Chen C, Zuo L, Heiner C, Burough F, Repetto M, Schlessinger D, D'Urso M (1996) Long-range sequence analysis in Xq28: thirteen known and six candidate genes in 219.4 kb of high GC DNA between the *RCP/GCP* loci. Hum Mol Genet 5:659-668

Collins FS (1995) Positional cloning moves from perditional to traditional. Nature Genet 9:347-350

Collins JE, Mungall AJ, Badcock KL, Fay JM, Dunham I (1997) The organization of the γ-glutamyl transferase genes and other low copy repeats in human chromosome 22q11. Genome Res 7:522-531

Colins JE, Cole CG, Smink LJ, Garrett CL, Leversha MA, Soderlund CA, Maslen GL, Everett LA, Rice KM, Coffey AJ, et al. (1995) A high-density YAC map of human chromosome 22. Nature 377:367-379

Conley ME, Beckwith JB, Mancer JFK, Tenckoff I (1979) The spectrum of DiGeorge syndrome. J Pediatr 94:883-890

Copeland NG, Jenkins NA, Gilbert DJ, Maltais LJ, Miller JC, Dietrich WF, Weaver A, Lincoln SE, Steen RG, Stein LD, Nadeau JH, Lander ES (1993) A genetic linkage map of the mouse: current applications and future prospects. Science 262:57-66

Courtay C, Heisterkamp N, Siest G, Groffen J (1994) Expression of multiple γ-glutamyltransferase genes in man. Biochem J 297:503-508

Crolla JA, Harvey JF, Sitch FL, Dennis NR (1995) Supernumerary marker 15 chromosomes: a clinical, molecular and FISH approach to diagnosis and prognosis. Hum Genet 95:161-170

Cross SH, and Bird AP (1995) CpG islands and genes. Curr Opin Genet Dev 5:309-314

Cuny G, Soriano P, Macaya G, Bernardi G (1981) The major components other mouse and human genome. I. Preparation, basic properties, and compositional heterogeneity. Eur J Biochem 115:227-233

Dallapiccola B, Pizzuti A, and Novelli G (1996) How many breaks do we need to CATCH on 22q11? Am J Hum Genet 59:7-11

D'Esposito M, Pilia G, Schlessinger D (1994) BLOCK-based PCR markers to find gene family members in human and comparative genome analysis. Hum Mol Genet 3:735-740

de Jong PJ, Yokobata K, Chen C, Lohman F, Pederson L, McNich J, Van Dilla M (1989) Human chromosome-specific partial digest libraries in lambda and cosmid vectors. Cytogenet Cell Genet 51:985

de la Chapelle A, Herva R, Koivisto M, Aulo O (1981) A deletion in chromosome 22 can cause DiGeorge syndrome. Hum Genet 57:253-256

Demczuk S, Thomas G, Aurias A (1996) Isolation of a novel gene from the DiGeorge syndrome critical region with homology to Drosophila gdl and to human LAMC1 genes. Hum Mol Genet 5:633-638

Demczuk S, Aledo R, Zucman J, Delattre O, Dsmaze C, Dauphinot L, Jalbert P, Rouleau GA, Thomas G, Aurias A (1995) Cloning of a balanced translocation breakpoint in the DiGeorge syndrome critical region and isolation of a novel potential adhesion receptor gene in its vicinity. Hum Mol Genet 4:551-558

Dietz HC, Cutting GR, Pyeritz RE, Maslen CL, Sakai LY, Corson GM, Puffenberger EG, Hamosh A, Nanthakumar EJ, Curristin SM, et al. (1991) Marfan syndrome caused by a recurrent de novo missense mutation in fibrillin gene. Nature 352:337-339

DiGeorge AM (1965) Discussion on a new concept of the cellular immunology. J Pediatr 67:907-908

Donlon TA, Lalande M, Wyman A, Latt A (1986) Isolation of molecular probes associated with the chromosome 15 instability in the Parader-Willi syndrome. Proc Natl Acad Sci 83:4408-4412

Dressler GR, Wilkinson JE, Rothenpieler UW, Patterson LT, Williams-Simons L, Westphal H (1993) Deregulation of Pax-2 expression in transgenic mice generates severe kidney abnormalities. Nature 362:65-67

Dressler GR, Deutsch U, Chowdhury K, Nornes HO, Gruss D (1990) Pax2, a new murine paired-box-containing gene and its expression in the developing excretory system. Development 109:787-795

Driscoll DA, Budarf ML, Emanuel BS (1992a) A genetic etiology for DiGeorge syndrome: consistent deletions and microdeletions of 22q11. Am J Hum Genet 50:924-933

Driscoll DA, Spinner NB, Budarf ML, McDonald-McGinn DM, Zackai EH, Golberg RB, Shprintzen RJ, et al. (1992b) Deletions and microdeletions of 22q11.2 in velo-cardio-facial syndrome. Am J Med Genet 44:261-268

Dutrillaux B (1973) Nouveau systeme de marquage chromosomique les bandes T. Chromosoma 41:395-402

Dworetzky SI, Boissard CG, Lum-Ragan JT, McKay MC, Post-Munson DJ, Trojanacki JT, Chang CD, Gribkoff VK (1996) Phenotypic alteration of a human BK (hSlo) channel by hSlo beta subunit coexpression : changes in blocker sensitivity, activation/relaxation, inactivation kinetics, and protein kinase modulation. J Neurosci 16:4543-4550

Eichler EE, Budarf ML, Rocchi M, Dearen LL, Doggett NA, Baldini A, Nelson DL, Mohrenweiser HW (1997) Interchromosomal duplications of the adrenoleukodystrophy locus: a phenomenon of pericentromeric plasticity. Hum Mol Genet 6:991-1002

Eikenboom JCJ, Vink T, Briet E, Sixma JJ, Reitsma PH (1994) Multiple substitutions in the von Willebrand factor gene that mimic the pseudogene sequence. Proc Natl Acad Sci USA 91:2221-2224

El-Shanti H, Hulseberg D, Murray JC, Patil SR (1993) A three generation minute supernumerary ring 22: association with cat-eye syndrome. Am J Hum Genet [suppl] 53:A126

Emanuel BS, Budarf ML, Sellinger B, Goldmuntz E, Driscoll DA (1992) Detection of midrodeletions of 22q11.2 with fluorescence in situ hybridization (FISH): diagnosis of DiGeorge syndrome (DGS), velo-cardio-facial (VCF) syndrome, CHARGE association and conotruncal cardiac malformations. Am J Hum Genet 51 [suppl] A3

Eppig JT, Nadeau JH (1995) Comparative maps: the mammalian jigsaw puzzel. Curr Opin Genet Dev 5:709-716

Epstein CJ (1990) The consequences of chromosome imbalance. Am J Med Genet [suppl] 7:31-37

Fields C, Adams MD, White O, Venter JC (1994) How many genes in the human genome? Nature Genet 7:345-346

Flint J, Thomas K, Micklem G, Raynham H, Clark K, Doggett NA, King A, Higgs DR (1997) The relationship between chromosome structure and function at a human telomeric region. Nature Genet 15:252-257

Foy C, Newton V, Wellesley D, Harris R, Read AP (1990) Assignment of the locus for Waardenberg syndrome type 1 to human chromosome 2q37 and possible homology to the Splotch mouse. Am J Hum Genet 46:1017-1023

Fraccaro M, Lindsten J, Ford CE, Iselius L , and 53 coauthors (1980) The 11q;22q translocation: a European collaborative analysis of 43 cases. Hum Genet 56:21-51

Friedman PA, Rao KW, Aylsworth AS (1987) Six patients with Townes-Brocks syndrome including five familial cases and an association with a specific inversion of chromosome 16. Am J Hum Genet [suppl] 41:A60

Frippiat J-P, Williams SC, Tomlinson IM, Cook GP, Cherif D, Paslier DL, Collins JE, Dunham I, Winter G, Lefranc M-P (1995) Organization of the human immunoglubin lambda light-chain locus on chromosome 22q11.2. Hum Mol Genet 4:983-991

Frohman MA, Dusk MK, Martin GR (1988) Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene specific oligonucleotide primer. Proc Natl Acad Sci USA 85:8998-9002

Galili N, Baldwin HS, Lund J, Reeves R, Gong W, Wang Z, Roe B, Emanuel BS, Nayak S, Mickanin C, Budarf ML, Buck CA (1997) A region of mouse chromosome 16 is syntenic to the DiGeorge, Velocardiofacial syndrome minimal critical region. Genome Research 7:17-26

Gardiner K (1995) Human genome organization. Curr Opi Genet Dev 5:315-322

Gerald PS, Davis C, Say B, Wilkins J (1972) Syndromal association of imperforate anus: the cat eye syndrome. Birth defects VIII, No 2:79-84

Gerald PS, Davis C, Say B, Wilkins J (1968) A novel chromosomal basis for imperforate anus (the "cat's eye" syndrome). Pediatr Res 2:297 (abstr)

Gluzman Y (1981) SV40-transformed simian cells support the replication of early SV40 mutants. Cell 23:175-182

Goldberg R, Marion R, Morderon W, Wiznia A, Shprintzen RJ (1985) Phenotypic overlap between cardio-velo-facial syndrome (VCF) and the DiGeorge sequence (DGS). Am J Hum Genet 37:A54

Goldmuntz E, Driscoll D, Budarf ML, Zackai EH, McDonald-McGinn DM, Biegel JA, Emanuel BS (1993) Microdeletions of chromosomal region 22q11 in patients with congenital conotruncal cardiac defects. J Med Gene 30:807-812

Gong W, Emanuel BS, Collins J, Kim DK, Wang Z, Chen F, Zhang G, Roe B, Budarf ML (1996) A transcriptional map of the DiGeorge and velo-cardio-facial syndrome minimal critical region on 22q11. Hum Mol Genet 5:789-800

Goulding MD, Chalepakis G, Deutsch U, Erselius JR,Gruss P (1991) Pax-3, a novel murine DNA binding protein expressed during early neurogenesis. EMBO J 10:1135-1147

Greenberg F, Elder FFB, Haffner P, Northrup H, Ledbetter DH (1988) Cytogenetic findings in a prospective series of patients with DiGeorge anomaly. Am J Hum Genet 43:605-611

Griffin CA, McKeon C, Israel MA, Gegonne A, Ghsdael J, Stehelin D, Douglass EC, et al. (1986) Comparison of constitutional and tumor-associated 11;22 translocations: non-identical breakpoints on chromosomes 11 and 22. Proc Natl Acad Sci USA 83:6122-6126

Groffen J, Stephenson JR, Hiesterkamp N, de Klein A, Bartram CR, Grosveld G (1984) Philadelphia chromosomal breakpoints are clustered within a limited region, bcr, on chromosome 22. Cell 36:93-99

Guanti G (1981) The aetiology of the cat eye syndrome reconsidered. J Med Genet 18:108-118

Guigo R, Knudsen S, Drake N, Smith TF (1992) Prediction of gene structure. J Mol Biol 226:51-60

Gustincich S, Manfioletti G, Del Sal G, Schneider C, Carnici P (1991). A fast method for high-quality genomic DNA extraction from whole human blood. Biotechniques 11:298-301

Gyapay G, Schmitt K, Fizames C, Jones H, Vega-Czarny N, Spillett D, Muselet D, Prud'Homme JF, Dib C, Auffray C, Morissette J, Goodfellow PN (1996) A rediation hybrid map of the human genome. Hum Mol Genet 5:339-346

Haab O (1878) Beitrage zu den angeoborenen Fahlern des Auges. Von Graefe's Arch Opthalmol 24:257-281

Haldane JBS, Sprunt AD, Haldane NM (1915) Reduplication in mice. J Genet 5:133-135

Halford S, Wilson DI, Daw SCM, Roberts C, Wadey R, Kamath S, Wickremasinghe A, Burn J, Mattei MG, Moormon AFM, Scambler PJ (1993) Isolation of a putative transcriptional regulator from the region of 22q11 deleted in DiGeorge syndrome, Shprintzen syndrome, and familial congenital heart disease. Hum Mol Genet 2:1577-1582

Haluska FG, Tsujimoto Y, Croce CM (1986) Oncogene activity by chromosome translocation in human malignancy. Ann Rev Genet 21:321-345

Harden DG, Klinger HP (1985) An International System for Human Cytogenetic Nomenclature (ISCN). S. Karger, Basel, New York.

Hassold TJ, Jacob PA (1984) Trisomy in man. Ann Rev Genet 18:69-97

Heiss NS, Rogner UC, Kioschis P, Korn B, Poustka A (1996) Transcriptional mapping in a 700-kb region around the DXS52 locus in Xq28: isolation of six novel transcripts and a novel ATPase isoform (hPMCA5). Genome Res 6:478-491

Heisterkamp N, Mulder MP, Langeveld A, Ten Hoeve J, Wang Z, Roe BA, Groffen J (1995) Localization of the human mitochondrial citrate transporter protein gene to chromosome 22q11 in the DiGeorge syndrome critical region. Genomics 29:451-456

Heisterkamp N, Groffen J (1988) Duplication of the bcr and gamma-glutamyl transpeptidase genes. Nucleic Acid Res 16:8045-8056

Hogan B, Beddington R, Costantini F, Lacy E (1996) Manipulating the mouse embryo, a laboratory manual, 2nd edition, Cold Spring Harbor Laboratory Press, CSH.

Holmes SE, Riazi MA, Weilong G, McDermid HE, Sellinger TS, et al. (1997) Disruption of the clathrin heavy chain-like gene (CLTCL) associated with features of DGS/VCFS: a balanced (21;22)(p12;q11) translocation. Hum Mol Genet 6: 357-367

Holmquist GP (1992) Review article: chromosome bands, their chromatin flavors, and their functional features. Am J Hum Genet 51:17-37

Homma K, Matsushita T, Natori S (1996) Purification, characterization, and cDNA cloning of a novel growth factor from the conditioned medium of NIH-Sape-4, an embryonic cell line of *Sarcophaga peregrina* (Flesh Fly). J Bio Chem 271:13770-13775

Hudson TJ, Stein LD, Gerety SS, Ma J, Castle AB, Silva J, Slonim DK, Baptista R, Kruglyak L, Xu SH, Hu X, et al. (1995) An STS-based map of the human genome. Science 270:1945-1954

Hulsebos TJM, Bijleveld EH, Riegman PHJ, Smink LJ, Dunham I (1996) Identification and characterization of NF1-related loci on human chromsomes 22, 14 and 2. Hum Genet 98:7-11

The Huntington's Disease Collaborative Research Group (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell 72:971-983

Ihle JN, Kerr IM (1995) Jaks and Sats in signaling by the cytokine receptor superfamily. Trends Genet 11:69-74.

Ioannou PA, Amemiya CT, Garnes J, Kroisel PM, Shizuya H, Chen C, Batzer MA, de Jong PJ (1994) A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. Nature Genet 6:84-89

Iwai N, Inagami T (1992) Identification of a candidate gene responsible for the high blood pressure of sponteneously hypersensitive rats. J Hyperten 10:1155-1157

Jaquez M, Driscoll DA, Li M, Emanuel BS, Hernandez I, Jaquez F, Lembert N, Ramirez J, and Matalon R (1997). Brief clinical report: unbalanced 15:22 translocation in a patient with both features of DiGeorge and velo-cardio-facial syndrome. Am J Med Genet, in press

Kaufman MH (1992) The Atlas of Mouse Development, Academic Press Inc.

Keller SA, Jones JM, Boyle A, Barrow LL, Killen PD, Green DG, Kapousta NV, Hitchcock PF, Swank RT, Meisler MH (1994) Kidney and retinal defects (Krd), a transgene-induced mutation with a deletion of mouse chromosome 19 that includes the Pax2 locus. Genomics 23:309-320

Kelley RI, Zackai EH, Emanuel BS, Kistenmacher M, Greenberg F, Punnett H (1982) The association of the DiGeorge anomaly with partial monosomy of chromosome 22. J Pediatr 101:197-200

Kennison JA (1995) The polycomb and trithorax group proteins of Drosophila: trans-regulators of homeotic gene function. Ann Rev Genet 29:289-303

Kim UJ, Shizuya H, Kang HL, Choi SS, Garrett C, Smink LJ, Birren BW, Korenberg JR, Dunham I, Simon MI (1996) A bacterial artificial chromosome-based framework contig map of human chromosome 22q. Proc Natl Acad Sci 93:6297-6301

Kinouchi A, Mori K, Ando M, Takao A (1976) Facial appearance of patients with conotruncal anomalies. Pediatr. Japan 17:84

Knoll JHM, Asamoah A, Pletcher BA, Wagstaff J (1995) Interstitial duplication of proximal 22q phenotypic overlap with Cat Eye Syndrome. Am J Med Genet 55:221-224

Krmpotic E, Rosnick MR, Zoller LM (1971) Genetic Counselling. Secondary nondisjunction in partial trisomy 13. Obstet Gynecol 37:381

Kuhn RM, Ludwig RA (1994) Complete sequence of the yeast artificial chromosome cloning vector pYAC4. Gene 141:125-127

Lamb BT, Gearhardt JD (1995) YAC transgenics and the study of genetics and human diseases. Curr Opin Genet Dev 5:342-348

Lammer EJ, and **Opitz** JM (1986). The DiGeorge anomaly as a developmental field defect. Am J Med Genet Suppl 2, 113-127

Lamour V, Le'Clouse Y, Desmaze C, Spetor M, Bodescot M, Aurias A, Osley MA, Lipinski M (1995) A human homolog of the S. cerevisae HIR1 and HIR2 transcriptional repressors cloned from the DiGeorge syndrome critical region. Hum Mol Genet 4:791-799

Lander ES, Botstein D (1986) Mapping complex genetic traits in humans: new methods using a complete RFLP linkage map. CSHSQB Vol L1:49-62

Larsen F, Gunderson G, Lopez R, Prydz H (1992) CpG islands as gene markers in the human genome. Genomics 13:1095-1107

Levy A, Demczuk S, Aurias A, Depetris D, Mattei MG, Philip N (1995) Interstitial 22q11 deletion excluding the ADU breakpoint in a patient with DGS. Hum Mol Genet 4:2417-2418

Lewin B (1994). Genes V. Oxford University Press and Cell Press

Li L, Krantz ID, Yu D, Genin A, Banta AB, Collins CC, Ming Q, Trask BJ, Kuo WL, Cochran J, Costa T, Pierpont MEM, Rand EB, Piccoli DA, Hood L, Spinner NB (1997) Alagille syndrome is caused by mutations in Human Jagged1, which encodes a ligand for Notch1. Nature Genet 16:243-250

Lindsay EA, Rizzu P, Antonacci R, Jurecic V, Delmas-Mata J, Lee CC Kim UJ, et al. (1996) A transcriptional map in the CATCH22 critical region: identification, mapping, and ordering of four novel transcripts expressed in heart. Genomics 32:104-112

Lipson AH, Yuille D, Angel M, Thompson PG, Vandervood JG, Becjenham EJ (1991) Velocardiofacial (Shprintzen) syndrome: an important syndrome for the dysmorphologist to recognize. J Med Genet 28:596-604

Litt M, White R (1985) A highly polymorphic locus in human DNA revealed by cosmid derived probes. Proc Natl Aca Sci USA 82:6206-6210

Liu S, Wong ML, Craik CS, Brodsky FM (1995) Regulation of clathrin assembly and trimerization defined using recombinant triskelion hubs. Cell 83:257-267

Long KR, Trofatter JA, Ramesh V, McCormick MK, Buckler AJ (1996) Cloning and characterization of a novel human clathrin heavy chain gene (CLTCL). Genomics 35:466-472

Maestre J, Tchenio T, Dhellin O, Heidmann T (1995) mRNA retroposition in human cells: processed pseudogene formation. EMBO J 14:6333-6338

Magenis RE, Sheehy RR, Brown MG, McDermid HE, White BN, Zonana J, and Weleber R (1988) Parental origin of the extra chromosome in the cat eye syndrome: evidence from heteromorphism and in situ hybridization analysis. Am J Med Genet 29:9-19

Makalowski W, Zhang J, Boguski MS (1996) Comparative analysis of 1196 orthologous mouse and human full-length mRNA and protein sequences. Genome Res 6:846-857

Mancuso DJ, Tuley EA, Westfield LA, Lester-Mancuso TL, Le Beau MM, Sorace JM, Sadler JE (1991) Human von Willebrand factor gene and pseudogene: structural analysis and differentiation by polymerase chain reaction. Biochemistry 30:253-269

Mannens M, Alders M, Redeker B, Bliek J, Steenman M, Wiesmeyer C, de Meulemeester M, Ryan A, Kalikin L, Voute T, De Kraker J, Hoovers J, Slater R, Feiberg A, Little P, Westerveld A (1996) Positional cloning of genes involved in the Beckwith-Wiedmann syndrome, hemihypertrophy, and associated childhood tumors. Med Pediat Oncol 27:490-494

Mascarello JT, Bastian JF, Jones MC (1989) Interstitial deletion of chromosome 22 in a patient with DiGeorge malformation sequence. Am J Med Genet 32:112-144

Martha A, Ferrell RE, Mintz-Hittner H, Lyons LA, Saunders GF (1994) Paired box mutations in familial and sporadic aniridia presicts truncated aniridia proteins. Am J Hum Genet 54:801-811

Mathews W, Jordan CT, Gavin M, Jenkins NA, Copeland NG, Lemishchka IR (1991) A receptor tyrosine kinase cDNA isolated from a population of enriched primitive hematopoietic cells and exhibiting close genetic linkage to c-kit. Proc Natl Acad Sci USA 88:9026-9030

McCarrey JR, Thomas K (1987) PGK gene lacks introns and possesses characteristics of a processed pseudogene. Nature 326:501-501

McDermid HE, Duncan AMV, Brasch JJA, Magenis HE, Sheely R, Burn J, Kardon N, Noel B, Schinzel A, Teshima I, and White BN (1986) Characterization of the supernumerary chromosome in cat eye syndrome. Science 232:646-648

McDermid HE, Budarf ML, Emanuel BS (1989) Toward a long-range map of human chromosomal band 22q11. Genomics 5:1-8

McDermid HE, Budarf ML, Emanuel BS (1993) Long-range map of human chromosome 22q11-22q12 between the Lambda immunoglobulin locus and Ewing sarcoma breakpoint. Genomics 18:308-318

McDermid HE, McTaggart KE, Riazi MA, Hudson TJ, Budarf ML, Emanuel BS, and Bell CJ (1996) Long-range mapping and construction of a YAC contig within the cat eye syndrome critical region. Genome Research 6:1149-1159

McDonnald-McGinn DM, Driscoll DA, Bason L, Christensen K, Lynch D, Sullivan K, Canning D, Zavod W, Quinn N, Rome J, Weinberg P, Clark BJ, Emanuel BS, Zackai EH (1995) Autosomal dominant "Optiz" GBBB syndrome due to 22q11.2 deletion. Am J Med Genet 59:103-113

McKusick VA (1991) Genomic mapping and how it has progressed. FASEB J 26:50-64

McManus OB, Helms LMH, Pallack L, Ganetzky B, Swanson R, Leonard RJ (1995). Functional role of the b subunit of high conductance calcium-activated potassium channels. Neuron 14:645-650

McPherson MJ, Quirke P, Taylor GR (1991) PCR, A Practical Approach, 1st ed. Oxford University Press

McTaggart KE, Budarf ML, Driscoll DA, Emanuel BS, Ferreira P, McDermid HE (1997). Cat eye syndrome chromosome breakpoints: identification of two intervals also associated with DiGeorge/Velocardiofacial syndrome deletion breakpoints. submitted to Am J Med Genet

Mears AJ, Duncan AMV, Budarf ML, Emanuel BS, Sellinger B, Siegel-Bartlet J, Greenberg CR, and McDermid HE (1994) Molecular characterization of the marker chromosome associated with cat eye syndrome. Am J Hum Genet 55:134-142

Mears AJ, El-Shanti H, Murray JC, McDermid HE, and Patil SR (1995) Minute supernumerary ring chromosome 22 associated cat eye syndrome: further delineation of the critical region. Am J Hum Genet 57:667-673

Mears AJ, PhD thesis (1995) Molecular characterization of cat eye syndrome. University of Alberta

Mezei LM, Storts DR (1994) in PCR technology: Current innovations, Griffin HG and Griffin AM eds. CRC Press, Boca Raton 21

Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, Liu Q, Cochran C, Bennett LM, Ding W, et al. (1994) A strong candidate for breast and ovarian cancer susceptibility gene BRCA1. Science 266:66-71

Morton NE (1991) Parameters of the human genome. Proc Natl Acad Sci 88:7474-7476

Mosser J, Sarde CO, Vicaire S, Yates JR, Mandel JL (1994) A new human gene (DXS1357E) with ubiquitous expression, located in Xq28 adjacent to the adrenoleukodystrophy gene. Genomics 22:469-471

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York, NY.

Nelson N (1991) Structure and pharmacology of the proton-ATPases. Trends Pharmacol Sci 12:71-75

Neufeld TP, Rubin GM (1994) The Drosophila peanut gene is required for cytokinesis and encodes a protein similar to yeast putative bud neck filament proteins. Cell 77:371-379

Noel B, Mottet J, Nantois Y, Quack B (1973) Identification of the small submetacentric supernumerary chromosome in the cat's-eye syndrome. J Hum Genet 21:23-32

North KN, Wu BL, Cao BN, Whiteman DAH, Korf BR (1995) CHARGE association in a child with *de novo* inverted duplication (14)(q22-q24.3). Am J Med Genet 57:610-614

Opitz JM, and Gilbert-Barness EF (1990) Reflections on the Pathogenesis of Down Syndrome. Am J Med Genet Supplement 7:38-51.

Osborne LR, Martindale D, Scherer SW, Shi XM, Huizenga J, Heng HHQ, Costa T, Pober B, Brinkman J, Rommens J, Koop B, Tsui LC (1996) Identification of genes from a 500-kb region at 7q11.23 that is commonly deleted in Williams syndrome patients. Genomics 36:328-336

Parimoo S, Patanjali SR, Shukla H, Chaplin DD, Weissman SM (1991) cDNA selection: efficient PCR approach for the selection of cDNAs encoded in large chromosomal DNA fragments. Proc Natl Acad Sci USA 88:9623-9627

Peifer, M (1993) Cancer, catenins, and cuticle pattern: a complex connection. Science 262:1667-1668

Pentao L, Wise CA, Chinault AC, Patel PI, Lupski JR (1992) Charcot-Marie-Tooth type 1A duplication appears to arise from recombination at repeat sequence flanking the 1.5 Mb monomer unit. Nature Genet 2:292-300

Pfeiffer RA, Heimann K, Heiming E (1970) Extra chromosome in "cat eye" syndrome. Lancet 2:297

Pizzuti A, Novelli G, Mari A, Ratti A, Colosimo A, amati F, Penso D, et al. (1996) Human homologue sequences to the Drosophila dishevelled segment-polarity gene are deleted in the DiGeorge syndrome. Am J hum Genet 58:722-729

Puffenberger EG, Hosodak K, Washington SS, Nakao K, deWit D, Yanagisawa M, Chakravarti A (1994) A missense mutation of the endothelin-B receptor gene in multigenic Hirschsprung's disease. Cell 79:1257-1266

Quinn TP, Peters KG, de Vries C, Ferrara N, Williams LT (1993) Fetal liver kinase 1 is a receptor for vascular endothelial growth factor and is selectively expressed in vascular endothelium. Proc Natl Aca Sci 90:7533-7537

Rahmani Z, Blouin JL, Creau-Goldberg N, Watkins PC, Mattei JF, Poissonier M, Prieur M, Chettouh Z, Nicole A, Aurias A, et al. (1990) Down syndrome critical region around D21S55 on proximal 21q22.3. Am J Med Genet [suppl] 7:98-103

Ramirez-Solis R, Liu P. Bradly A (1995) Chromosome engineering in mice. Nature 378:720-724

Regnier V, Meddeb M, Lecointre G, Richard F, Duverger A, Nguyen VC, Dutrillaux B, Berheim A, Danglot G (1997) Emergence and scattering of multiple neurofibromatosis (NF1)-related sequences during hominoid evolution suggest a process of pericentromeric interchromosomal transposition. Hum Mol Genet 6:9-16

Reiss JA, Weleber RG, Brown MG, Bangs CD, Loverin EW, Magenis RE (1985) Tandem duplication of proximal 22q: a cause of cat eye syndrome. Am J Med Genet 20:165-171

Rizzu P, Baldini A (1994) Subchromosomal band interval mapping and ordering of DNA markers in the region 3q26.3-q27 involved in the dup (3q) syndrome. Genomics 24:580-582

Robinson MS (1994) The role of clathrin, adaptors and dynamin in endocytosis. Curr Opin Cell Biol 4:538-544

Rommens JM, Lin B, Hutchinson GB, Andrews SE, Goldberg YP, Glaves ML, Graham R, Lai V, McArthur J, Nasir J, et al. (1993) A transcription map of the region containing the Huntington disease gene. Hum Mol Genet 2:901-907

Rommens JM, Iannuzzi MC, Kerem BS, Drumm ML, Melmer G, Dean M, Rozmahel R, Cole JL, Kennedy D, Hidaka N, Zsiga M, Buchwald M, Riordan JR, Tsui L-C, Collins FS (1989) Identification of the cystic fibrosis gene: chromosome walking and jumping. Science 245:1059-1065

Ross ME, Allen KM, Srivastava AK, Featherstone T, Gleeson JG, Hirsch B, Harding BN, Andermann E, Abdullah R, Berg M, Czapansky-Bielman D, Flanders DJ, Guerrini R, Motte J, Mira AP, Scheffer I, Berkovic S, Scaravilli F, King RA, Ledbetter DH, Schlessinger D, Dobyns WB, Walsh CA (1997) Linkage and physical mapping of X-linked lissencephaly/SBH (XLIS): a gene causing causing neuronal migration defects in human brain. Hum Mol Genet 6:555-562

Rouleau GA, Haines JL, Bazanowski A, Colella-Crowley A, Troffater JA, Wexler NS, Conneally PM, Gusella JF (1989) A genetic linkage map of the long arm of human chromosome 22. Genomics 4:1-6

Saccone S, Caccio S, Kusuda J, Andreozzi L, Bernardi G (1996) Identification of the gene-richest bands in human chromosomes. Gene 174:85-94

Saccone S, De Sario A, Wiegant J, Raap AK, Della Valle G, Bernardi G (1993) Correlations between isochores and chromosomal bands in the human genome. Proc Natl Aca Sci USA 90:11929-11933

Samani NJ, Whitemore SA, Kaiser MA, Harris J, See CG, Gallen DF, Lodwick D (1994) Chromosomal assignment of the human SA gene to 16p13.11 and demonstration of its expression in the kidney. Biochem Biophys Res Cummun 199:862-868

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular Cloning: a Laboratory Manual, second edition, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY)

Sanyanusin P, Schimmenti LA, McNoe LA, Ward TA, Pierpont MEM, Sullivan MJ, Dobyns WB, and Eccles MR (1995) Mutation of the PAX2 gene in a family with optic nerve colobomas, renal anomalies, and vesicoureteral reflux. Nat Genet 9:358-364

Scambler PJ, Carey AH, Wyse RKH, Roach S, Dumanski JP, Nordenskjold M, Williamson R (1991) Microdeletions within 22q11 associated with sporadic and familial DiGeorge syndrome. Genomics 10:201-206

Schachenmann G, Schmid W, Fraccaro M, Mannini A, Tiepolo L, Perona GP, Sartori E (1965). Chromosomes in coloboma and anal atresia. Lancet 2:290

Schinzel AA, Basaran S, Karaman B, Yuksel-Apak M, Robinson WP (1994) Maternal uniparental disomy 22 has no impact on the phenotype. Am J Hum Genet 54:21-24

Schinzel A (1991) tetrasomy 12p (Pallister-Killian syndrome). J Med Genet 28:122-125

Schinzel A, Schmid W, Fraccaro M, Tiepolo L, Zuffardi O, Opitz JM, Lindsten J, Zetterquist P, Enell H, et al. (1981) The "cat-eye-syndrome": Dicentric small marker chromosome probably derived from a No. 22 (tetrasomy 22 pter-q11) associated with a characteristic phenotype. Hum Genet 57:148-158

Schmickel RD, Knoller M (1977) Characterization and localization of the human genes for ribosomal ribonucleic acid. Pediat Res 11:929-935

Schnur RE, Knowlton RG, Musarella MA, Muenke M, Nussbaum RL (1990) Partial deletions of a sequence family ("DXS278") and its physical linkage to steroid sulfatase as detected by pulsed-field gel electrophoresis. Genomics 8:255-262

Schreck RR, Breg WR, Erlanger BF, Miller OJ (1977) Preferential derivation of abnormal human G-group-like chromosomes from chromosome 15. Hum Genet 36:1-12

Schuler DG, Boguski MS, Stewart EA, Stein LD, Gyapay G, Rice K, White RE, Rodriguez-Tome P, Aggarwal A, Bajorek , et al. (1996) A gene map of the human genome. Science 274:540-546

Schwartz DC, and Cantor CR (1984) Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. Cell 37:67-75

Serville F, Lacombe D, Saura R, Billeaud C, Sergent MP (1993) Townes-Brocks syndrome in an infant with translocation t(5;16). Genet Counseling 4:109-112

Shapiro BL (1983) Down Syndrome-a disruption of homeostasis. Am J Med Genet 14:241-269

Shapiro BL (1989) The pathogenesis of aneuploid phenotypes: the fallacy of explanatory reductionism. Am J Med Genet 33:146-158

Sharkey AM, McLaren L, Carroll M, Fantes J, Green D, Wilson D, Scambler PJ, Evans HJ (1992) Isolation of anonymous DNA markers for human chromosome 22q11 from a flow-sorted library, and mapping using hybrids from patients with DiGeorge syndrome. Hum Genet 89:73-78

Shimmenti LA, Cunliffe HE, McNoe LA, Ward TA, French MC, Shim HH, Zhang YH, et al. (1997) Further delineation of renal-coloboma syndrome in patients with extreme variability of phenotype and identical PAX2 mutations. Am J Hum Genet 60:869-878

Shindoh N, Kudoh J, Maeda H, Yamaki A, Minoshima S, Shimizu Y, Shimizu N (1996) Cloning of a human of the Drosophila minibrain/rat Dyrk gene from "the Down syndrome critical region" of chromosome 21. Biochem Biophys Res Commun 225:92-99

Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y, Simon M (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Eschericia coli* using an F-factor-based vector. Proc Natl Acad Sci 89:8794-8797

Shprintzen RJ, Goldberg RB, Lewin ML, Sidoti EJ, Berkman MD, Argamaso RV, Young D (1978) A new syndrome involving cleft palate, cardiac anomalies, typical facies, and learning disabilities: Velo-cardio-facial syndrome. Cleft Palate J 15:56

Shprintzen RJ, Wang F, Goldberg R, Marion R (1985) The expanded velo-cardio-facial syndrome: additional features of the most common clefting syndrome. Am J Hum Genet 37:A77

Sinclair AH, Berta P, Palmer MS, Hawkins JR, Griffith BL, Smith MJ, Foster JW, Frischauf AM, Lovell-Badge R, Goodfellow PN (1990) A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature 346:240-244

Sirvastava S, Zou ZQ, Pirollo K, Blattner W, Chang EH (1990) Germ-lone transmission of a mutated p53 gene in a cancer-prone family with Li-Fraumeni syndrome. Nature 348:747-749

Smith DJ, Stevens ME, Sudanagunta SP, Bronson RT, Makhinson M, Watabe AM, O'Dell TJ, Fung J, Weier HU, Cheng JF, Rubin EM (1997) Functional screening of 2 Mb of human chromosome 21q22.2 in transgenic mice implicates minibrain in learning defects associated with Down syndrome. Nature Genet 16:28-36

Solovyev VV, Salamov AA, Lawrence CB (1994) Predicting internal exons by oligonucleotide composition and discriminant analysis of spliceable open residing frames. Nucleic Acids Res 22:5156-5163

Sossin WS, Kreiner T, Barinaga M, Schilling J, Scheller RH (1989) A dense core vesicle protein is restricted to the cortex of granules in the exocrine atrial gland of Aplysia california. J Bio Chem 264:16933-16940

Stedman's Medical Dictionary (1982). 24 ed., Waverly Press, Inc.

Sternberg N (1990) Bacteriophage P1 cloning system for the isolation, amplification, and recovery of DNA fragments as large as 100 kilobase pairs. Proc Natl Acad Sci USA 87:103-107

Stevens CA, Carey JC, Shigeoka AO (1990) DiGeorge anomaly and velo-cardio-facial syndrome. Pediatrics 85:526-530

Strong WB (1968) Familial syndrome of right-sided aortic arch, mental deficiency and facial dysmorphism. J Pediatr 73:882-888

Summer FB (1922) Linkage in peromyscus. Am Nat 56:412-417

Sutherland HF, Wadey R, McKie JM, Taylor C, Atif U, Johnstone KA, Halford S, Kim UJ, Goodship J, Baldini A, Scambler PJ (1996) Identification of a novel transcript disrupted by a balanced translocation associated with DiGeorge syndrome. Am J Hum Genet 59:23-31

Sutherland GR, Baker E, Hyland VJ, Gallen DF, Stahl J, Gough NM (1989) The gene for human leukemia inhibitory factor (LIF) maps to 22q12. Leukemia 3:9-13

Takahata N, Satta Y (1997) Evolution of primate lineage leading to modern humans: phylogenetic and demographic inferences from DNA sequences. Proc Natl Acad Sci 94:4811-4815

Tassabehji M, Read AP, Newton VE, Harris R, Balling R, Gruss P, Strachan T (1992) Waardenberg's syndrome patients have mutations in the human homologue of the Pax-3 paired box gene. Nature 355:635-636

Tazi J, and Bird A (1990) Alternative chromatin structure at CpG islands. Cell 60:909-920

Terman BI, Carrion ME, Kovacs E, Rasmussen BA, Eddy RL, Shows TB (1991) Identification of a new endothelial cell growth factor receptor tyrosine kinase. Oncogene 6:1677-1683

Tejedor F, Zhu XR, Kaltenback E, Akermann A, Baumann A, Canal I, Heisenberg M, Fischbach KF, Pongos O (1995) Minibrain: a new protein kinase family involved in postembryonic neurogenesis in Drosophila. Neuron 14:287-301

Timmerman V, Nelis E, Van Hul W, Nieuwenhuijsen BW, Chen KL, Wang S, Ben Othman K, Leach RJ, Hanemann CO, et al. (1992) The peripheral myelin protein gene PMP-22 is contained within the Charcot-Marie-Tooth disease type 1A duplication. Nat Genet 1:171-175

Toomey KE, Mohandas T, Leisti J, Szalay G, Kaback MM (1977) Further delineation of the supernumerary chromosome in the Cat-Eye syndrome. Clin Genet 12:275-284

Terman BI, Dougher-Vermazen M, Carrion ME, Dimitrov D, Armellino DC, Gospodarowicz D, Bohlen P (1992) Identification of the KDR as a receptor for vascular endothelial cell growth factor. Biochem Biophys Res Commun 187:1579-1586

Townes PL, Brocks ER (1972) Hereditary syndrome of the imperforate anus with hand, foot, and ear anomaly. J Pediatr 81:321-326

Trofatter JA, Long KR, Murrell JR, Stotler CJ, Gusella JF, Buckler AJ (1995) An expression-independent catalog of genes from human chromosome 22. Genome Res 5:214-224

Trofatter JA, MacCollin MM, Rutter JL, Murrell JR, Parry MP, Eldridge DM, Kley N, Menon AG, Pulaski K, Haase V, Ambrose C, Munroe D, Bove D, Haines JL, Martuza RL, MacDonald ME, Siezinger BR, Short MP, Buckler AJ, Gusella JF (1993) A novel meosin-, ezrin-, radixin-like gene is a candidate for the neurofibromatosis 2 tumor suppressor. Cell 72:791-800

Urioste M, Visedo G, Sanchis A, Sentis C, Villa A, Ludena P, Hortiguela JL, Martinez-Frias ML, Fernandez-Piqueras J (1994) Dynamic mosaicism involving an unstable supernumerary der(22) chromosome in cat eye syndrome. Am J Med Genet 49:77-82

Valdes JM, Tagle DA, Collins FS (1994) Island rescue PCR: a rapid and efficient method for isolating transcribed sequences from yeast artificial chromosomes and cosmids. Proc Natl Acad Sci USA 91:5377-5381

Van Dyke DL, Weiss M, Pai GS (1977) The origin and behavior of two isodicentric bisatellited chromosomes. Am J Hum Genet 29:294-300

Waardenberg PJ (1951) A new syndrome combining developmental anomalies of the eyelids, eyebrows, and nose root with pigmentary defects of the iris and head hair and congenital deafness. Am J Hum Genet 3:195-253

Wadey R, Daw S, Taylor C, Atif U, Kamath S, Halford S, O'Donnell H, Wilson D, Goodship J, Burn J, Scambler P (1995) Isolation of a gene encoding an integral membrane protein from the vicinity of a balanced translocation breakpoint associated with DiGeorge syndrome. Hum Mol Genet 4:1027-1033

Walknowska J, Peakman D, Weleber RG (1997) Cytogenetic investigation of cat-eye syndrome. Am J Opthalmol 84:477-486

Watkin-Chow DE, Buckwalter MS, Newhouse MM, Lossie AC, Brinkmeier ML, Camper SA (1997) Genetic mapping of 21 genes on mouse chromosome 11 reveals duplications in linkage conservation with human chromosome 5. Genomics 40:114-122

Webb T (1994) Inv dup (15) supernumerary marker chromosomes. J Med Genet 31:585-594

Weleber RG, Walknowska J, Peakman D (1977) Cytogenetic investigation of cat-eye syndrome. Am J Optalmol 84:477-486

Whang-Peng J, Triche TJ, Knutsen T, Miser J, Douglass EC, Israel MA (1984) Chromosome translocation in peripheral neuroepithelioma. N Engl J Med 311:584-585

Wilkinson DG, Bailes JA, McMahon AP (1987) Expression of the proto-oncogene int-1 is restricted to specific neural cells in the developing mouse embryo. Cell 50:79-88

Wilson DI, Burn J, Scambler PJ, Goodship (1993) Syndrome of the month: DiGeorge syndrome, part of CATCH-22. J Med Genet 30:852-856

Wisniewski L, Hassold T, Heffelfinger J, Higgins JV (1979) Cytogentic and clinical studies in five cases of inv dup(15). Hum Genet 50:259-300

Wolff D, and Schwartz S (1992) Characterization of Robertsonian translocation by using fluorescence in situ hybridization. Am J Hum Genet 50:174-181

Worley KC, Wiese BA, Smith RF (1995) BEAUTY: An enhanced BLAST-based search tool that integrates multiple biological information resources into sequence similarity search results. Genome Res 5:173-184

Xu Y, Einstein JR, Mural RJ, Shah M, Uberbacher EC (1994) An improved system for exon recognition and gene modeling in human DNA sequences. IN "ISMB-94 Proceedings Second International Conference on Intelligent Systems for Molecular Biology (R. Altman, D Brutlag, P Karp, R Lathrop, and D Searls, Eds), pp. 376-384, AAAI Press, Menlo Park.

Yaspo ML, Gellen L, Mott R, Korn B, Nizetic D, Poustka AM, Lehrach H (1995) Model for a transcript map of human chromosome 21: isolation of new coding sequences from exon and enriched cDNA libraries. Hum Mol Genet 4:1291-1304

Yen PH, Li XM, Tsai SP, Johnson C, Mohandas T, Shapiro LJ (1990) Frequent deletions of the human X chromosome distal short arm result from recombination between low copy repetitive elements. Cell 61:603-610

Young D, Shprintzen RJ, Goldberg RB (1980) Cardiac malformations in the velo-cardio-facial syndrome. Am J Cardiol 46:43-48

Zackai EH, Emanuel BS (1980) Site-specific reciprocal translocation, t(11;22)(q23;q11) in several unrelated families with 3:1 meiotic disjunction. Am J Med Genet 7:507-521

Zoubak S, Clay O, Bernardi G (1996) The gene distribution of the human genome. Gene 174:95-102

Appendix

The sequence of the following exons was determined.

KCNMBLP:

e1-
CATTCGGAGAGAAGAATCAACCTGCACTGCCATCCACACAGATATCATGGACG
ACTGGCTGGACTGTGCCACCTGTGGTGTGACACTGCCGTGGTCACGGGAAGTA
CTCGTGTCTTCGGGTGTTTGTGAACCTCACCCATTCAGCTCAGAAAGCTCTCCT
ACATTATAATGAAGAGGCTGTCCAGATAAATCCCAAG

SAHL1:

e11-
TGGTGAAGGCATTGATTGTCCTGACCCCACAGTTTCTGTCCCATGACAAGGACC
AGCTGACTAAGGAGCTGCAGCAGCATGTAAAGTCATTCACAGCCCCAT

e22-
AGGTAACCCAGAGAACACAGCCGAAGTGGAATGTGGGGACGTCTGCGACACC
GGGGACAGAGCAACTGTGGATGAAGAGGGCTACATCTGGTTCCTGGGGAGGA
TCATGACACTATCAATGCCTCTGG

a-
CGCCGAGGCCTGCCACCTCCTCCACCAACAAACGGAAGCCACCACATCTTCTC
CGAGACTCGCTCTGACCACGCCGCCGATG

b-
AGTTTTAAATTTAATTTAATTTAATTTTCATAGTTACAGAGTCTTGTTATGTTGCC
CAGGCTGGTGTCAAACTCCTGGTTTCAAGGGATCCTCTCTCCTTGGCCC

d-
AAACTGGATACCTTAGGATTTTCTGTATACAAGATCATGACATTTGGGAGTAAG
ACAGTTTTACTTTTTCCTTTCTAATCTGGATCCTTTTATATATCAATCGCCTAATC
ATCCTGGTCAGAACCAAGGCTGCATAGACCTGCTGAACATGGCCATCCTTGTCT
TGTTCCCAATCTTAGGGGAAAGCATTCAGTCTTGCACCATTAAATCTGTTAACTG
TTAGGTTTTCACAGATGCTTTTCATCATGCTGATGAAGTTCCCTCCTGTTTCTAG
TTTGTTGAGGGTTTTTTTCCATGCATGGGTAGATAATTTCACTCTCATTTACAGA
AAAGGAAACTGTCATTCAGGCAGTCTGCCTTCAAAGATCAAGCCCTGGATCACA
GTGCTATCTCTACGCACTCTTCCATTTAATTTTCCCTCTGGCCCTCCTCCGAATT
CTCTCTTTCAGCTCACTTGGGAATGTTGGTGTTCTCCAGGAATCTCAGCTCTTTG
CTCATGTGCTTCTC

e-
ATTTCATATTCCCACTGGCAGTGTATGAGGGTTTCGATTTCTCCACCTCCTCTCC
AGCACTTGGCGTTATCTGTCTTGATTGGTTGTATTCATT

IDGFL:

e5-
TCCCGTGGAGTCTCCATCTGAGCCCTTTCCCTAGTCCAGGCATCCCGATGTTGG
TGGATGGCCCATCTGAGCGGCCAGCCCTGTGCTTCTTGCTGTTGGCTGTGGCA
ATGTCTTTCTTCGGCTCAGCTCTATCCATAAATGAAACACGGGCGCATCTGTTG
TTGAAAGAAAAAAATGATGCGGGTGGGGGGCGGCTGGTGCTGAACACCAAG
GAGGAGCTGGCCAATGAGAGGCTCATGACGCTCAAAATCGCTGAGATGAAGG
AGGCCATGAGGACCCTGATATTCCCACCCAGCATGCACTTTTTCCAGGCCAA

Other exons:

e6-
GGATCACCAGAATTCTGGAGCTCGAGCGGCCGCTGCAGGATCACGCGTGCGG
CGTCCGGCGTCTCTTCCGAAAGCGTGGTCACCGCATCGAAACCGAGGAAGGA
GAAACAGACAATCGTCGCCCCG

e7-
GTTTCTGGATACCAGACCAGGCAGAAGGCAGCACAGACCATGCACCATGCAGA
TGTGGCCTCGAGACCCATCATCCACCTCAAGTCCTGACTCGGCACTGGCAGGT
CAGCATCTCAGGCCCTGATCAAG

e8-
CCACACCTTTGATTCCAGCTCCTCCCAACACCCCATCTCCACGCAGAGTCTCAC
CAACACAGACTCAGAAGACAGCGGAGAGAGGTCTCTTTTCCCGCAGAACCCGG
CATCTGCATTTCCTGTTTCCGGTGGAACCAGCAGTTCAGTCCCGCCGAGGAGC
ACTGGTAACATCCACTACGCGGTCCTGGACTTCCAGCCGAGCCAAG

CES-TK-
ATCAGGTGGACGGCCCTGGATGTCTCTGTGCAGCCCATGCCTCTAGCTTGGAC
CATTCACCGCCCAGGTACCTCCTCGCTAGGTCTCCAAGCCCCACTGGTCATGTT
AAAAGCCATGGGGCAGGGCTGGGATGGGTTCAGGCTCCATGTTCTTCCTGAGA
AGGCTGAAGGGTTGCATCCCTCCAAACAGCCCAAGGCATAGTATCCACTGCTA
CCCCACTCAGTCTCGGTGAACAAGCTTGTCGGGATGTTTGGTCCTTTGGCGTA